ISCRAM 2025
HALIFAX. NOVA SCOTIA. CANADA
MAY 18-21

Dynamic Fusion of Large Language Models for Crisis Communication

Xiaoying Song, Anirban Saha Anik, Vanessa Frías-Martínez, Lingzi Hong

UNT UNIVERSITY OF NORTH TEXAS

UNIVERSITY OF MARYLAND

Work-in-Progress Paper

Presented By Anirban Saha Anik

Outline

- Motivation
- Research Question
- Methods
- Results
- Takeaways

Motivation & Problem

"Real help requests posted during Hurricane Irma on Twitter (X)"

Anonymous @anonymous

Help! Rescue Needed. 35 weeks pregnant. Waist deep water. 2815 34th St. Port Arthur TX 77640, Her name is Haley Dav......

#RescueNeeded #PregnantWoman #PortArthur #FloodRescue

4:24 PM - 12 September 2017

2 13 4

Anonymous @anonymous

WE NEED A RESCUE TEAM HERE
A residence roof falling in please help!
5801 N. Houston Rosslyn apt 1712
(1) 6 month old baby inside!

#HoustonFlood #HurricaneRelief #EmergencyRescue #HoustonTX #FloodRescue #USCG

12:32 AM - 02 October 2017

1 44 2

Anonymous @anonymous

@USCG WE NEED A RESCUE TEAM HERE ASAP!
Small child + 2 adults
Contact Name : Ices Bragg +1 409-995-####

#RescueNeeded #HurricaneRelief #FloodRescue

9:45 PM - 07 September 2017

1 44 2

Source: Hurricane Irma Tweets *(Hong et al. 2020)*

Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, *9*(1), 29.

Motivation & Problem

"Real help requests posted during Hurricane Irma on Twitter (X)"

Anonymous @anonymous

Help! $\boxed{\text{Rescue Needed. 35 weeks pregnant.}}$ Waist deep water. 2815 34th St. Port Arthur TX 77640, Her name is Haley Dav......

#RescueNeeded #PregnantWoman #PortArthur #FloodRescue

4:24 PM - 12 September 2017

2 13 4

Anonymous @anonymous

WE NEED A RESCUE TEAM HERE
A residence roof falling in please help!
5801 N. Houston Rosslyn apt 1712
(1) 6 month old baby inside!

#HoustonFlood #HurricaneRelief #EmergencyRescue #HoustonTX #FloodRescue #USCG

12:32 AM - 02 October 2017

1 44 2

Anonymous @anonymous

@USCG WE NEED A RESCUE TEAM HERE ASAP!
Small child + 2 adults
Contact Name : Ices Bragg +1 409-995-####

#RescueNeeded #HurricaneRelief #FloodRescue

9:45 PM - 07 September 2017

1 44 2

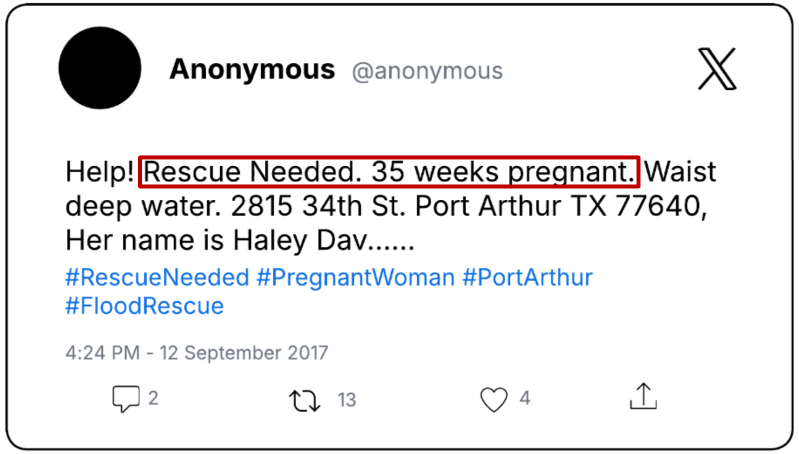
Medical Emergency

Source: Hurricane Irma Tweets *(Hong et al. 2020)*

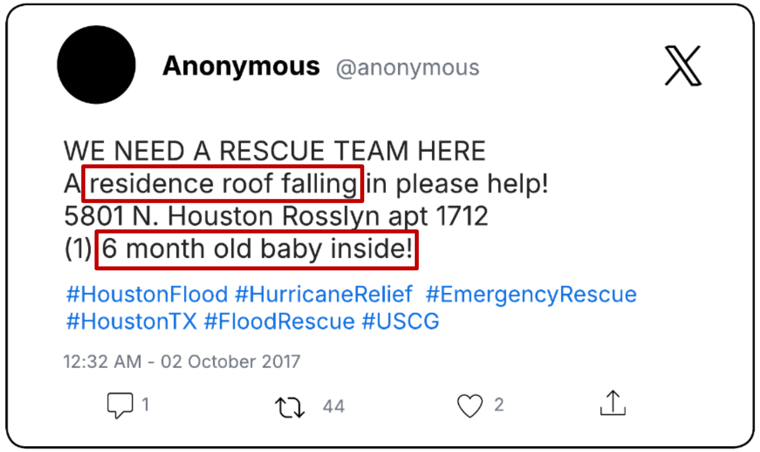
Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, *9*(1), 29.

Motivation & Problem

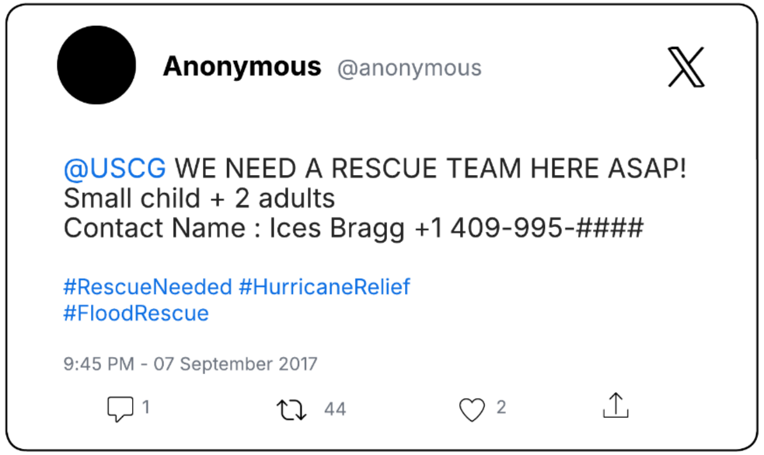
"Real help requests posted during Hurricane Irma on Twitter (X)"



Medical Emergency



Structural Collapse

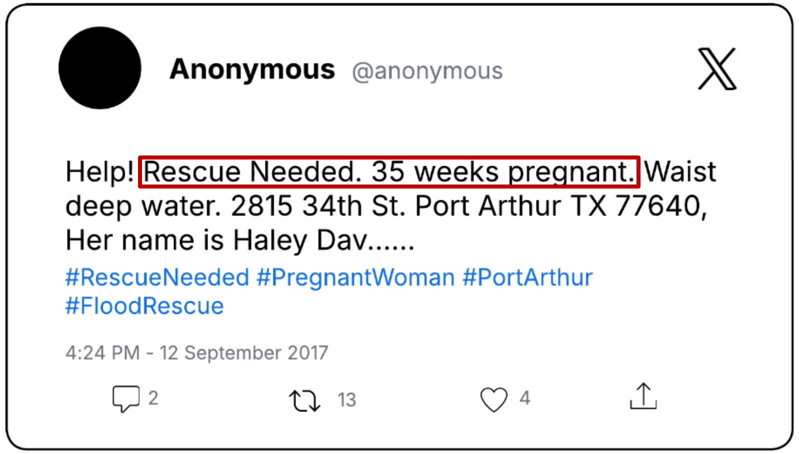


Source: Hurricane Irma Tweets *(Hong et al. 2020)*

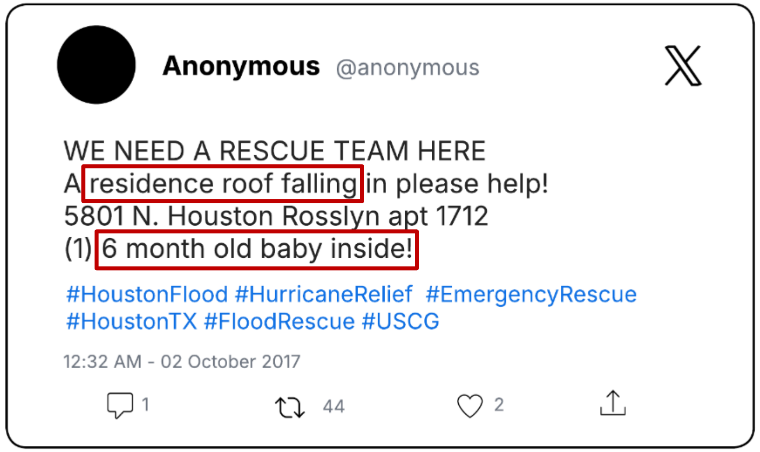
Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, *9*(1), 29.

Motivation & Problem

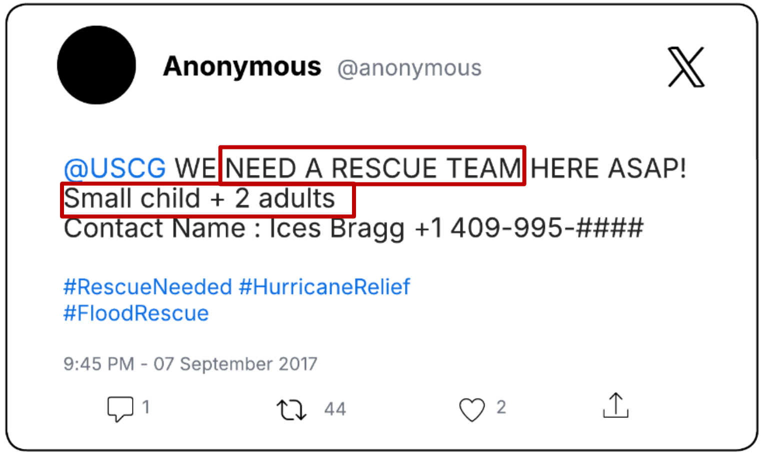
“Real help requests posted during Hurricane Irma on Twitter (X)”



Medical Emergency



Structural Collapse



Family in Danger

Source: Hurricane Irma Tweets *(Hong et al. 2020)*

Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, *9*(1), 29.

Motivation & Problem

"Real help requests posted during Hurricane Irma on Twitter (X)"

Anonymous @anonymous

Help! Rescue Needed. 35 weeks pregnant. Waist deep water. 2815 34th St. Port Arthur TX 77640, Her name is Haley Dav......

#RescueNeeded #PregnantWoman #PortArthur #FloodRescue

4:24 PM - 12 September 2017

2 13 4

Medical Emergency

Anonymous @anonymous

WE NEED A RESCUE TEAM HERE
A residence roof falling in please help!
5801 N. Houston Rosslyn apt 1712
(1) 6 month old baby inside!

#HoustonFlood #HurricaneRelief #EmergencyRescue #HoustonTX #FloodRescue #USCG

12:32 AM - 02 October 2017

1 44 2

Structural Collapse

Anonymous @anonymous

@USCG WE NEED A RESCUE TEAM HERE ASAP!
Small child + 2 adults
Contact Name : Ices Bragg +1 409-995-####

#RescueNeeded #HurricaneRelief #FloodRescue

9:45 PM - 07 September 2017

1 44 2

Family in Danger

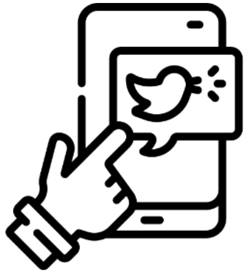
"Thousands of people post help requests like these—but many receive no response."

Source: Hurricane Irma Tweets *(Hong et al. 2020)*

Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, *9*(1), 29.

During Disasters, People Turn to Social Media for Help...

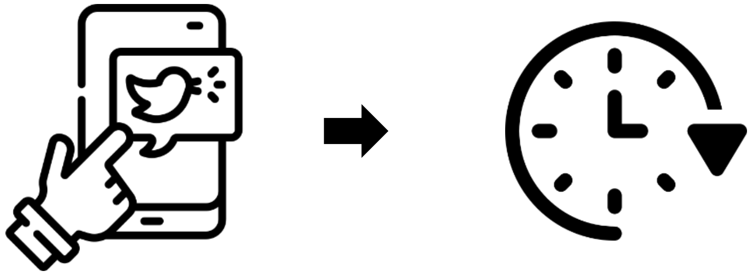
- People post urgent needs on Twitter (X), hoping someone will respond.



Tweet posted

During Disasters, People Turn to Social Media for Help…

- People post urgent needs on Twitter (X), hoping someone will respond.
- Emergency responders can't keep up with the **volume and urgency** of social media requests.



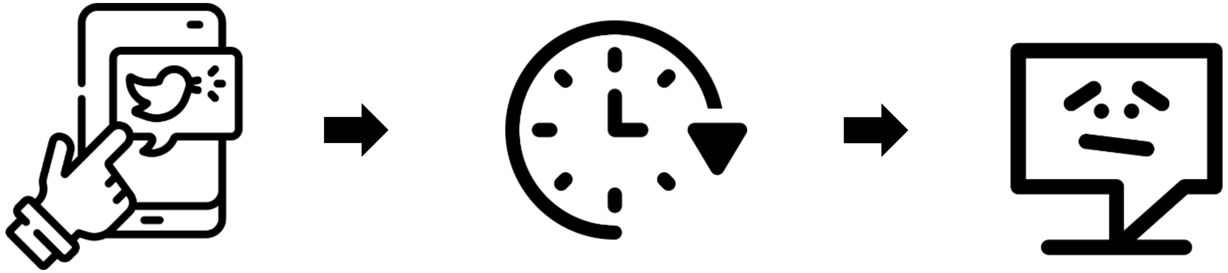
Tweet posted **Hours pass**

During Disasters, People Turn to Social Media for Help…

- People post urgent needs on Twitter (X), hoping someone will respond.
- Emergency responders can’t keep up with the **volume and urgency** of social media requests.
- Responses are often missing, delayed, or inaccurate.

During Disasters, People Turn to Social Media for Help...

- People post urgent needs on Twitter (X), hoping someone will respond.
- Emergency responders can't keep up with the **volume and urgency** of social media requests.
- Responses are often missing, delayed, or inaccurate.



Tweet posted → **Hours pass** → **No response**

What if AI could step in and help respond—fast, helpful, and reliable?

Our Goal

Main Objective

Build an AI assistant that can generate fast, helpful, and human-like responses to crisis-related social media posts.

Our Goal

Main Objective

Build an AI assistant that can generate fast, helpful, and human-like responses to crisis-related social media posts.

What We Aim to Achieve

- **Understand the help-seeker's intent** from short, noisy tweets
- **Provide factually accurate and grounded information** using trusted sources
- **Deliver clear, actionable steps** people can follow during an emergency
- **Show empathy** to build trust and provide emotional support
- **Balance all of these in real-time responses**

Our Goal

Main Objective

Build an AI assistant that can generate fast, helpful, and human-like responses to crisis-related social media posts.

What We Aim to Achieve

- **Understand the help-seeker's intent** from short, noisy tweets
- **Provide factually accurate and grounded information** using trusted sources
- **Deliver clear, actionable steps** people can follow during an emergency
- **Show empathy** to build trust and provide emotional support
- **Balance all of these in real-time responses**

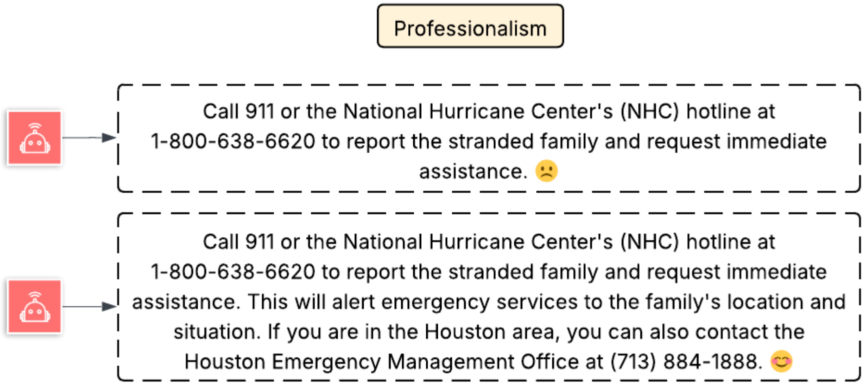
Our Approach

We explore **three methods** using Large Language Models (LLMs):

1. **Instruction Prompt** – guided generation
2. **RAG** – Document-grounded generation
3. **Dynamic Fusion** – *Our proposed method to **combine the best of both***

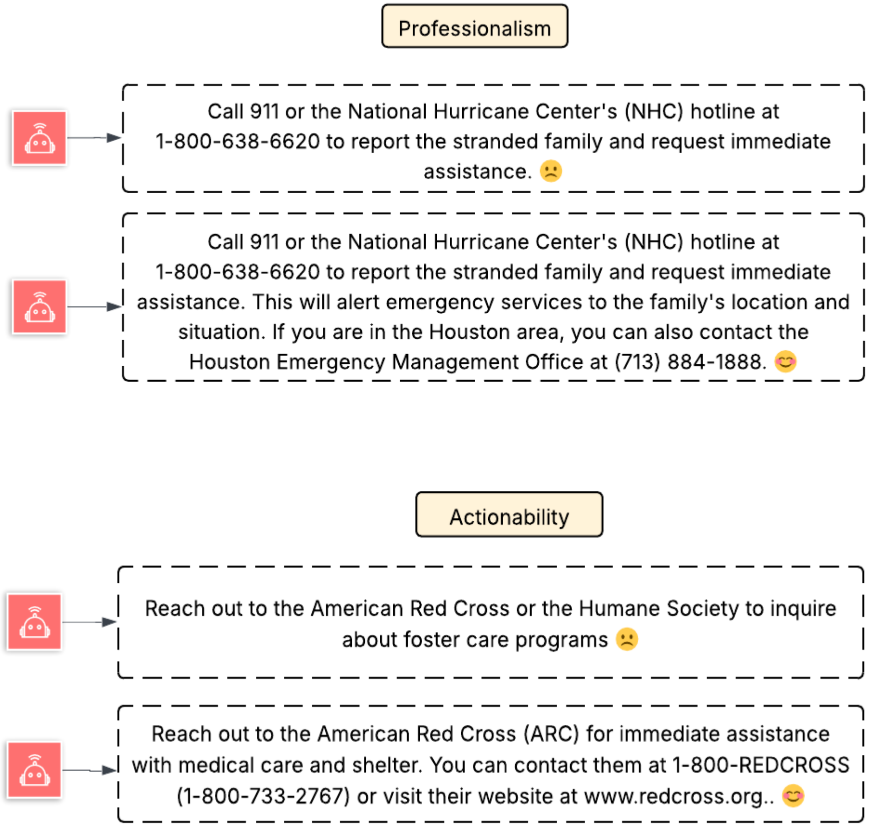
Evaluation Dimensions

- **Professionalism** – The response sounds official, reliable, and well-informed.



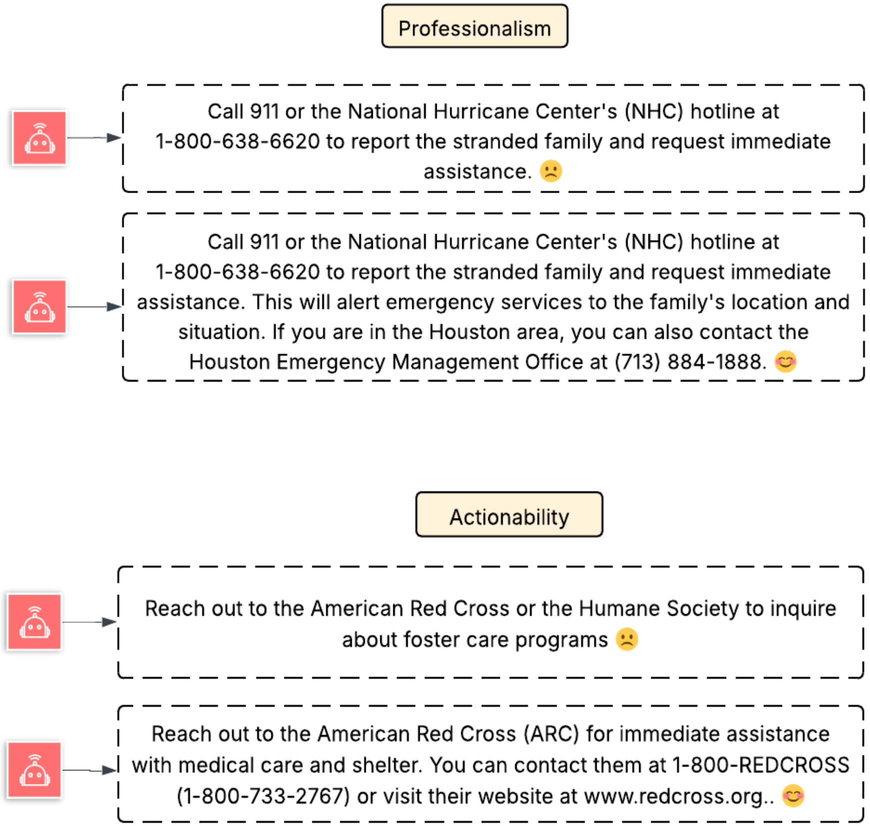
Evaluation Dimensions

- **Professionalism** – The response sounds official, reliable, and well-informed.
- **Actionability** – The response gives clear, step-by-step guidance that people can follow during a crisis.



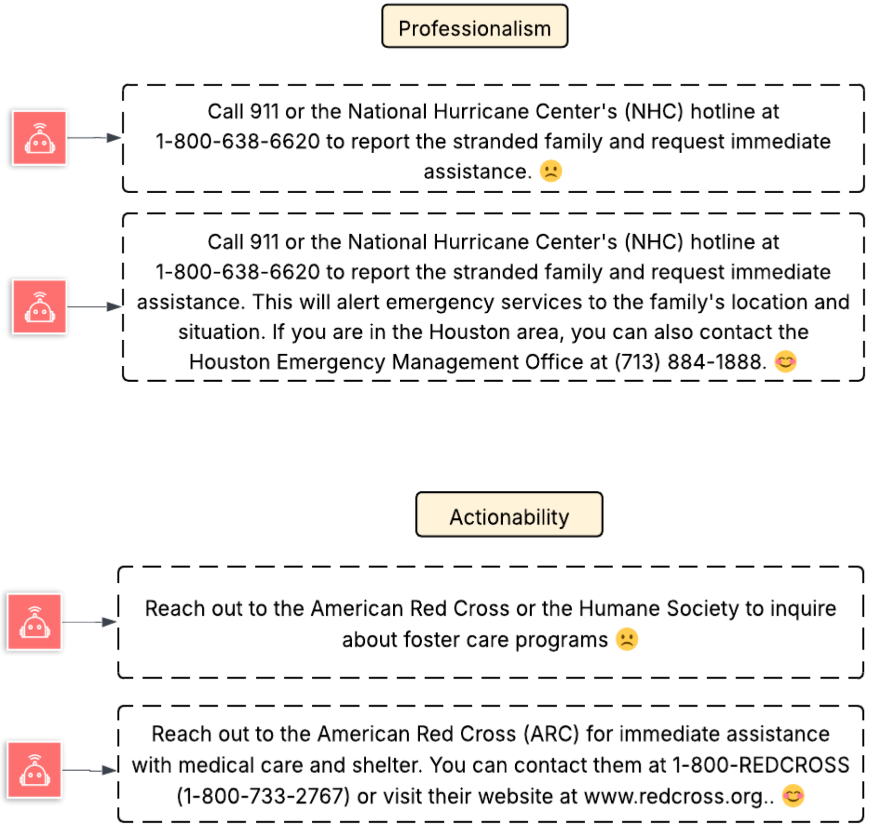
Evaluation Dimensions

- **Professionalism** – The response sounds official, reliable, and well-informed.
- **Actionability** – The response gives clear, step-by-step guidance that people can follow during a crisis.
- **Empathy** – The response is compassionate, supportive, and emotionally appropriate.



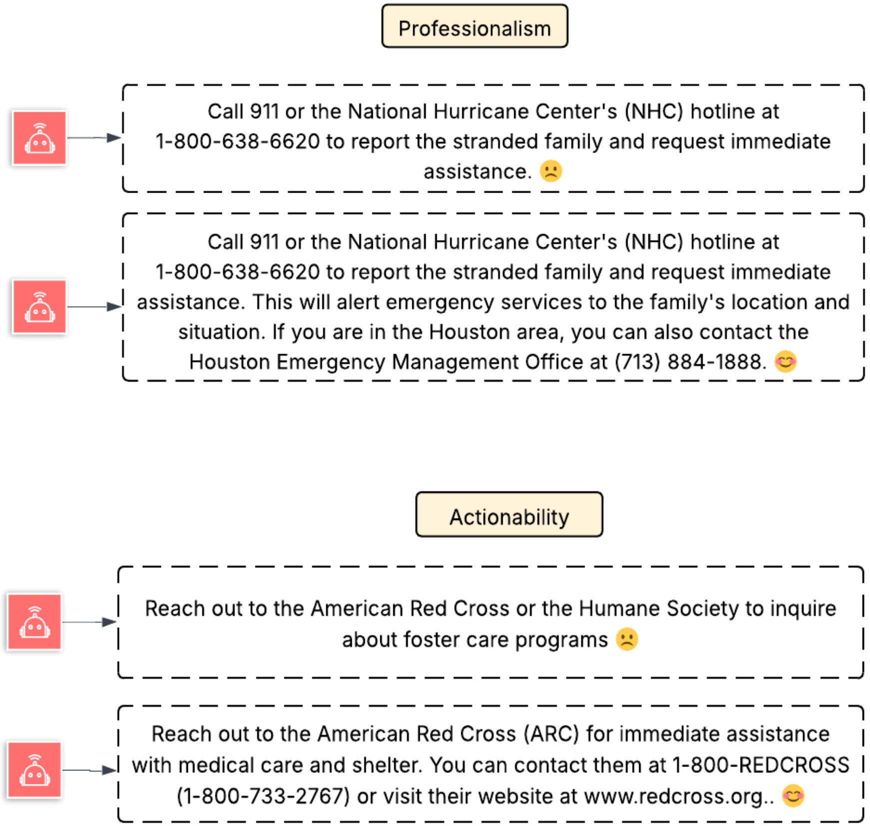
Evaluation Dimensions

- **Professionalism** – The response sounds official, reliable, and well-informed.
- **Actionability** – The response gives clear, step-by-step guidance that people can follow during a crisis.
- **Empathy** – The response is compassionate, supportive, and emotionally appropriate.
- **Relevance** – The response directly addresses the specific needs or situation described in the post.



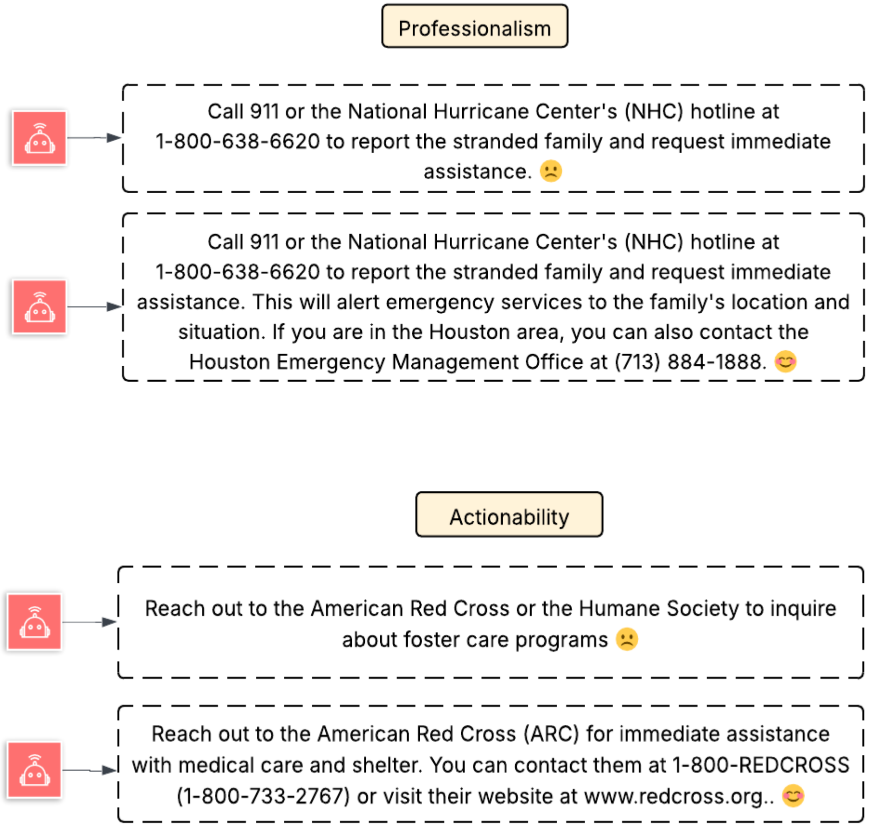
Evaluation Dimensions

- **Professionalism** – The response sounds official, reliable, and well-informed.
- **Actionability** – The response gives clear, step-by-step guidance that people can follow during a crisis.
- **Empathy** – The response is compassionate, supportive, and emotionally appropriate.
- **Relevance** – The response directly addresses the specific needs or situation described in the post.
- **Overall Quality Score** – We calculate a weighted score **Q** = 0.4 × Professionalism + 0.4 × Actionability + 0.1 × Empathy + 0.1 × Relevance



Evaluation Dimensions

- **Professionalism** – The response sounds official, reliable, and well-informed.
- **Actionability** – The response gives clear, step-by-step guidance that people can follow during a crisis.
- **Empathy** – The response is compassionate, supportive, and emotionally appropriate.
- **Relevance** – The response directly addresses the specific needs or situation described in the post.
- **Overall Quality Score** – We calculate a weighted score
  Q = 0.4 × Professionalism + 0.4 × Actionability + 0.1 × Empathy + 0.1 × Relevance

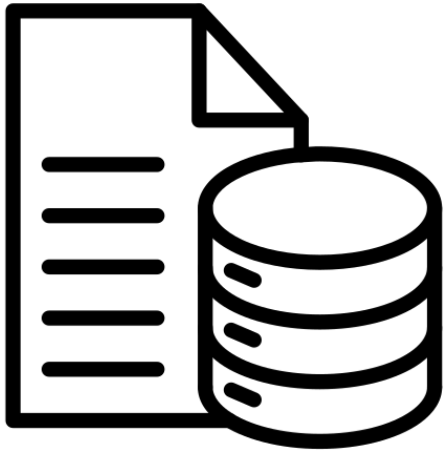


"Good crisis communication must be clear, credible, caring, and on point."

Dataset

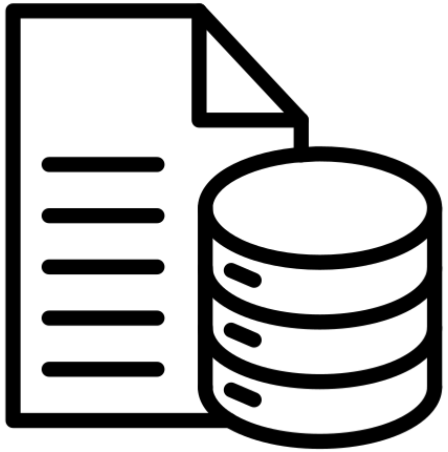
Raw Dataset

- **Source:** Hurricane Irma Tweets *(Hong et al. 2020)*[1]
- **Timeframe:** August 15 – October 12, 2017
- **Location:** 6 Southern U.S. states
  (Florida, Georgia, South Carolina, North Carolina, Tennessee, Alabama)
- **Raw Data:** 1,013,313 tweets from 127,181 users



[1]Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, *9*(1), 29.

Dataset



Raw Dataset

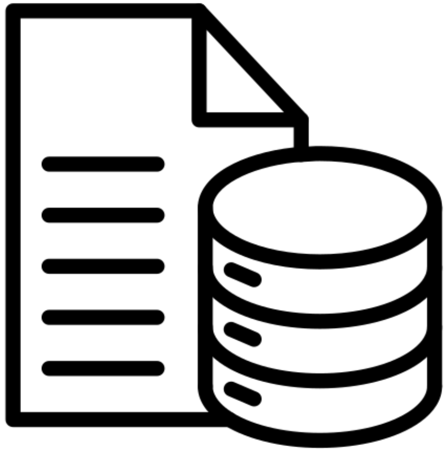
- **Source:** Hurricane Irma Tweets *(Hong et al. 2020)*[1]
- **Timeframe:** August 15 – October 12, 2017
- **Location:** 6 Southern U.S. states
  (Florida, Georgia, South Carolina, North Carolina, Tennessee, Alabama)
- **Raw Data:** 1,013,313 tweets from 127,181 users

Detecting Needs-Related Tweets

- Trained **three RoBERTa models** on crisis-labeled datasets *(Alam et al., 2021a; Alam et al., 2021b)*[2,3]
  - HumanAID, CrisisBench, Turkey Earthquake Relief
- Tweets labeled as **"needs-related"** if all three models agree

[1] Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, *9*(1), 29.
[2] Alam, F., Sajjad, H., Imran, M., & Ofli, F. (2021, May). CrisisBench: Benchmarking crisis-related social media datasets for humanitarian information processing. In Proceedings of the International AAAI conference on web and social media (Vol. 15, pp. 923-932).
[3] Alam, F., Qazi, U., Imran, M., & Ofli, F. (2021, May). Humaid: Human-annotated disaster incidents data from twitter with deep learning benchmarks. In Proceedings of the International AAAI Conference on Web and social media (Vol. 15, pp. 933-942).

Dataset



Raw Dataset

- **Source:** Hurricane Irma Tweets *(Hong et al. 2020)*[1]
- **Timeframe:** August 15 – October 12, 2017
- **Location:** 6 Southern U.S. states
  (Florida, Georgia, South Carolina, North Carolina, Tennessee, Alabama)
- **Raw Data:** 1,013,313 tweets from 127,181 users

Detecting Needs-Related Tweets

- Trained **three RoBERTa models** on crisis-labeled datasets *(Alam et al., 2021a; Alam et al., 2021b)*[2,3]
  - HumanAID, CrisisBench, Turkey Earthquake Relief
- Tweets labeled as **"needs-related"** if all three models agree

Final Labeled Dataset

- **540 verified needs-related tweets**
- Used as input for LLM-based response generation

[1]Hong, L., & Frias-Martinez, V. (2020). Modeling and predicting evacuation flows during hurricane Irma. *EPJ Data Science*, *9*(1), 29.
[2]Alam, F., Sajjad, H., Imran, M., & Ofli, F. (2021, May). CrisisBench: Benchmarking crisis-related social media datasets for humanitarian information processing. In Proceedings of the International AAAI conference on web and social media (Vol. 15, pp. 923-932).
[3]Alam, F., Qazi, U., Imran, M., & Ofli, F. (2021, May). Humaid: Human-annotated disaster incidents data from twitter with deep learning benchmarks. In Proceedings of the International AAAI Conference on Web and social media (Vol. 15, pp. 933-942).

Method 1 – Instruction Prompt

What is Instruction Prompting?

Give an LLM clear, structured instructions to generate a response for crisis-related tweets.

Prompt Design

> You are an AI assistant designed to provide professional, actionable, empathetic, and relevant advice for someone seeking help related to a hurricane on social media.
>
> Given the following tweet expressing needs during a hurricane, provide a detailed solution. If you don’t know the answer, clearly state, "I don’t know".
>
> <u>Guidelines:</u>
>
> - Prioritize immediate actions, labeled as Step 1, Step 2, etc.
> - For each step, add a short explanation.
> - Include links, organizations, or contact info if relevant.

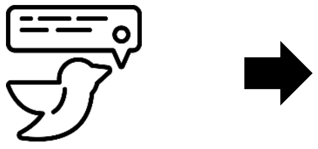
Method 1 – Instruction Prompt

What is Instruction Prompting?

Give an LLM clear, structured instructions to generate a response for crisis-related tweets.

Prompt Design

> You are an AI assistant designed to provide professional, actionable, empathetic, and relevant advice for someone seeking help related to a hurricane on social media.
>
> Given the following tweet expressing needs during a hurricane, provide a detailed solution. If you don’t know the answer, clearly state, "I don’t know".
>
> Guidelines:
>
> - Prioritize immediate actions, labeled as Step 1, Step 2, etc.
> - For each step, add a short explanation.
> - Include links, organizations, or contact info if relevant.



Tweet

Method 1 – Instruction Prompt

What is Instruction Prompting?

Give an LLM clear, structured instructions to generate a response for crisis-related tweets.

Prompt Design

> You are an AI assistant designed to provide professional, actionable, empathetic, and relevant advice for someone seeking help related to a hurricane on social media.
>
> Given the following tweet expressing needs during a hurricane, provide a detailed solution. If you don't know the answer, clearly state, "I don't know".
>
> Guidelines:
>
> - Prioritize immediate actions, labeled as Step 1, Step 2, etc.
> - For each step, add a short explanation.
> - Include links, organizations, or contact info if relevant.

Tweet

Instruction Prompt

Method 1 – Instruction Prompt

What is Instruction Prompting?

Give an LLM clear, structured instructions to generate a response for crisis-related tweets.

Prompt Design

You are an AI assistant designed to provide professional, actionable, empathetic, and relevant advice for someone seeking help related to a hurricane on social media.

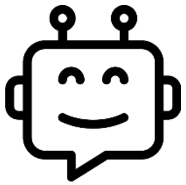
Given the following tweet expressing needs during a hurricane, provide a detailed solution. If you don’t know the answer, clearly state, "I don’t know".

Guidelines:

- Prioritize immediate actions, labeled as Step 1, Step 2, etc.
- For each step, add a short explanation.
- Include links, organizations, or contact info if relevant.

Tweet

Instruction Prompt



Response

Method 2 – RAG (Retrieval-Augmented Generation)

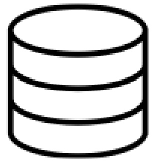
What is RAG?
Combine the power of LLMs with **trusted external documents** (e.g., FEMA guides) to ground the response in **real-world knowledge**.

Method 2 – RAG (Retrieval-Augmented Generation)

What is RAG?
Combine the power of LLMs with **trusted external documents** (e.g., FEMA guides) to ground the response in **real-world knowledge**.

How It Works – 3 Steps

- **Knowledge Base Construction**
  - We curated FEMA documents[1] (e.g., Individual Assistance Guide).
  - Split into small chunks for retrieval.



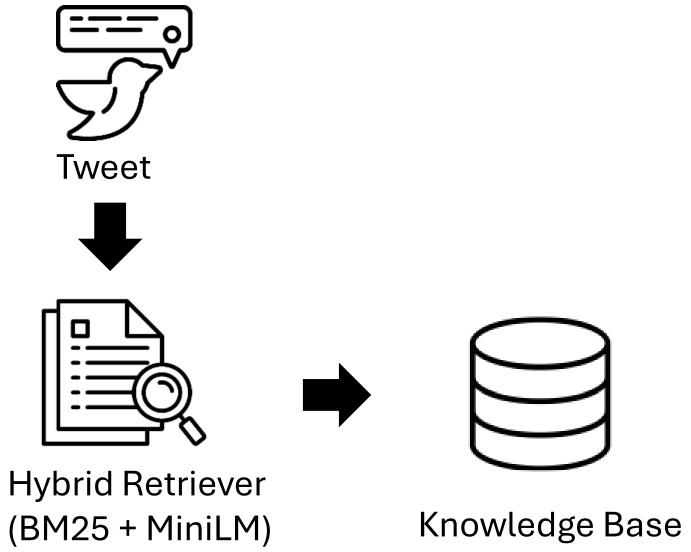
[1] https://www.fema.gov/

Method 2 – RAG (Retrieval-Augmented Generation)

What is RAG?
Combine the power of LLMs with **trusted external documents** (e.g., FEMA guides) to ground the response in **real-world knowledge**.

How It Works – 3 Steps

- **Knowledge Base Construction**
    - We curated FEMA documents[1] (e.g., Individual Assistance Guide).
    - Split into small chunks for retrieval.
- **Document Retrieval**
    - Hybrid retriever = **Keyword-based (BM25) + Semantic search (MiniLM)**
    - Top-5 relevant documents selected.



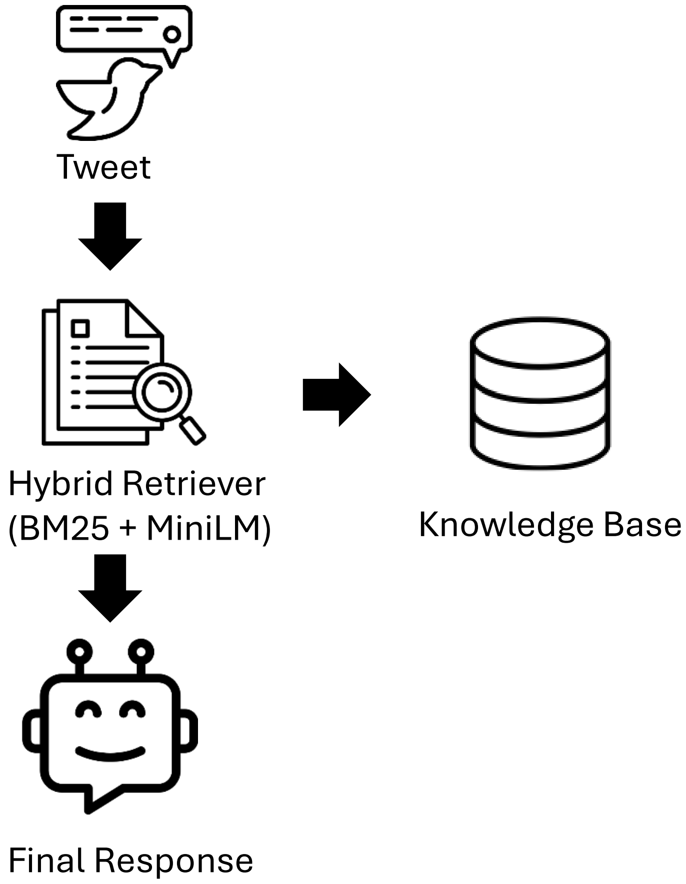
[1]https://www.fema.gov/

Method 2 – RAG (Retrieval-Augmented Generation)

What is RAG?
Combine the power of LLMs with **trusted external documents** (e.g., FEMA guides) to ground the response in **real-world knowledge**.

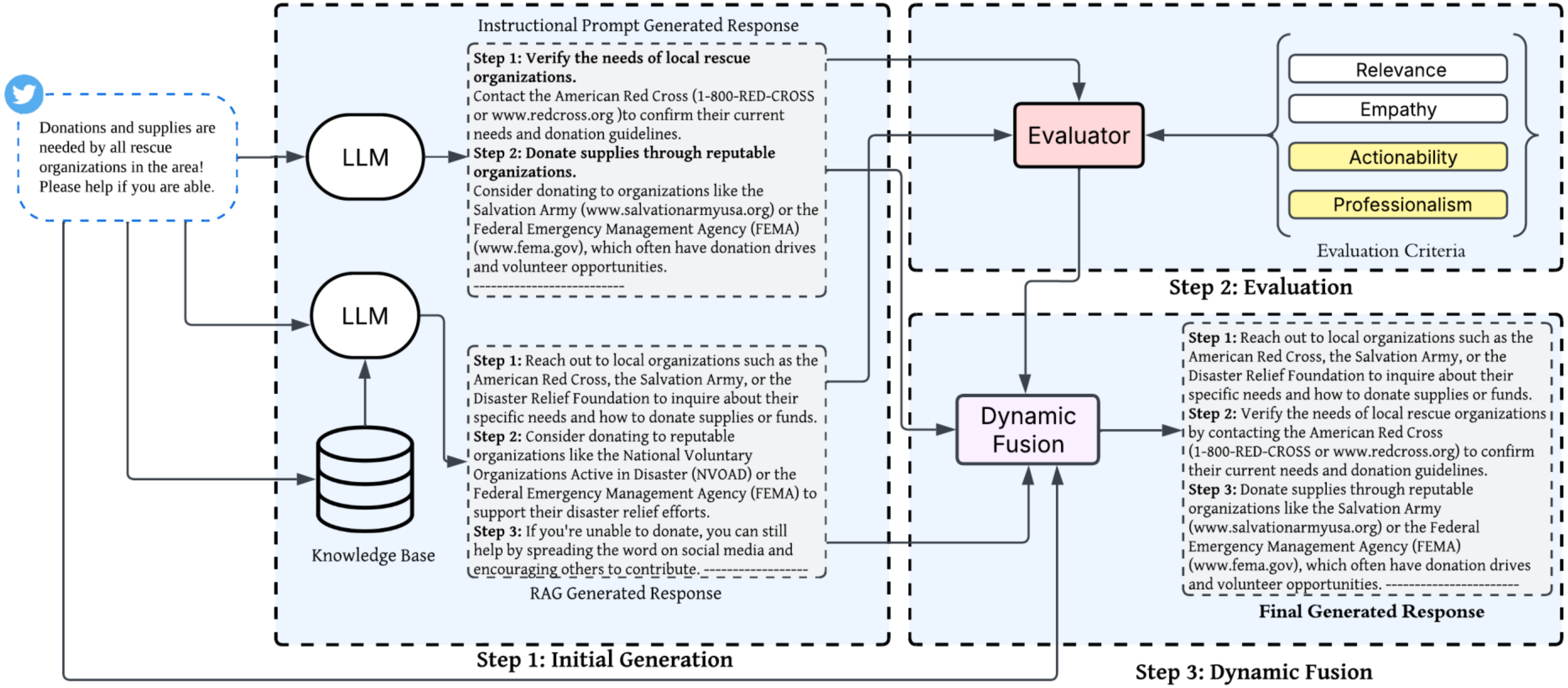
How It Works – 3 Steps

- **Knowledge Base Construction**
  - We curated FEMA documents[1] (e.g., Individual Assistance Guide).
  - Split into small chunks for retrieval.
- **Document Retrieval**
  - Hybrid retriever = **Keyword-based (BM25)** + **Semantic search (MiniLM)**
  - Top-5 relevant documents selected.
- **Response Generation**
  - The retrieved documents are **concatenated with the user tweet** and fed into the LLM.
  - Generates a grounded, informative reply.

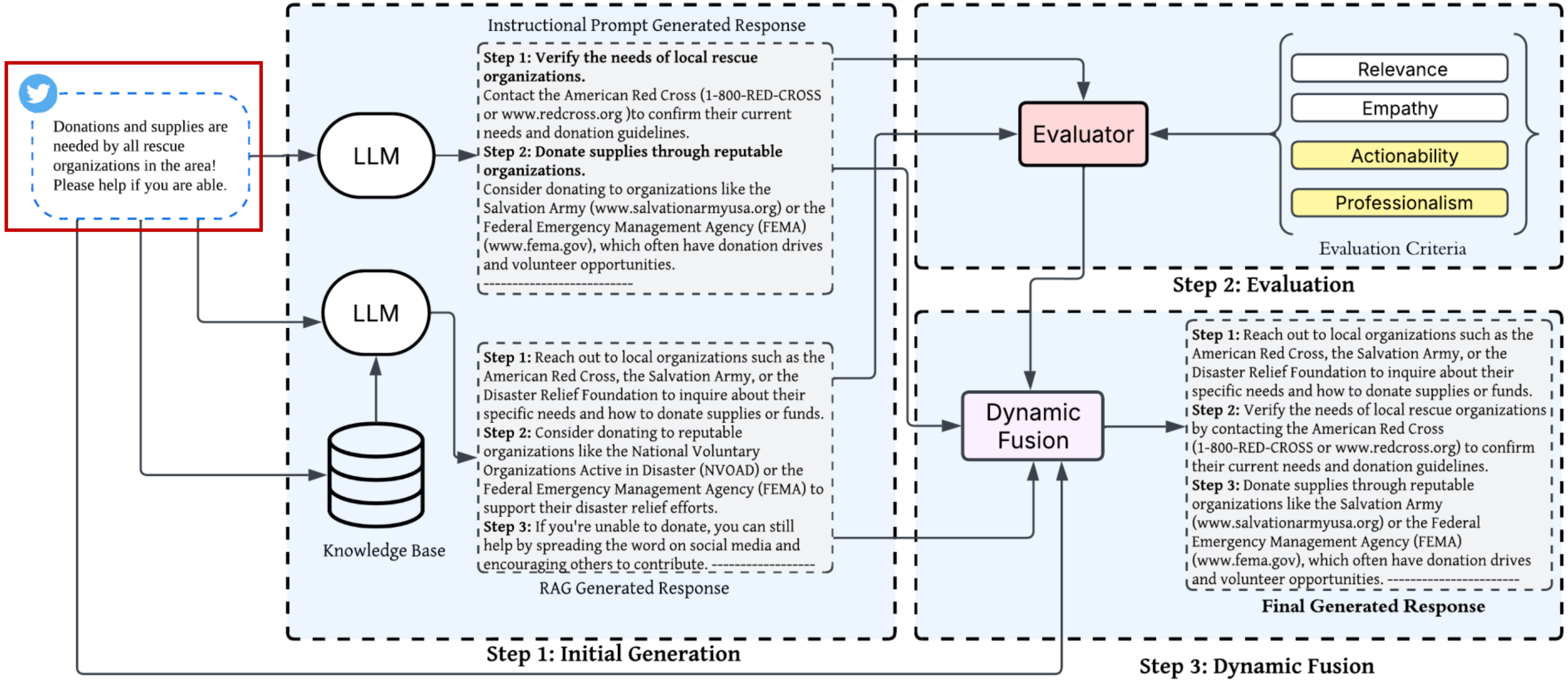


[1]https://www.fema.gov/

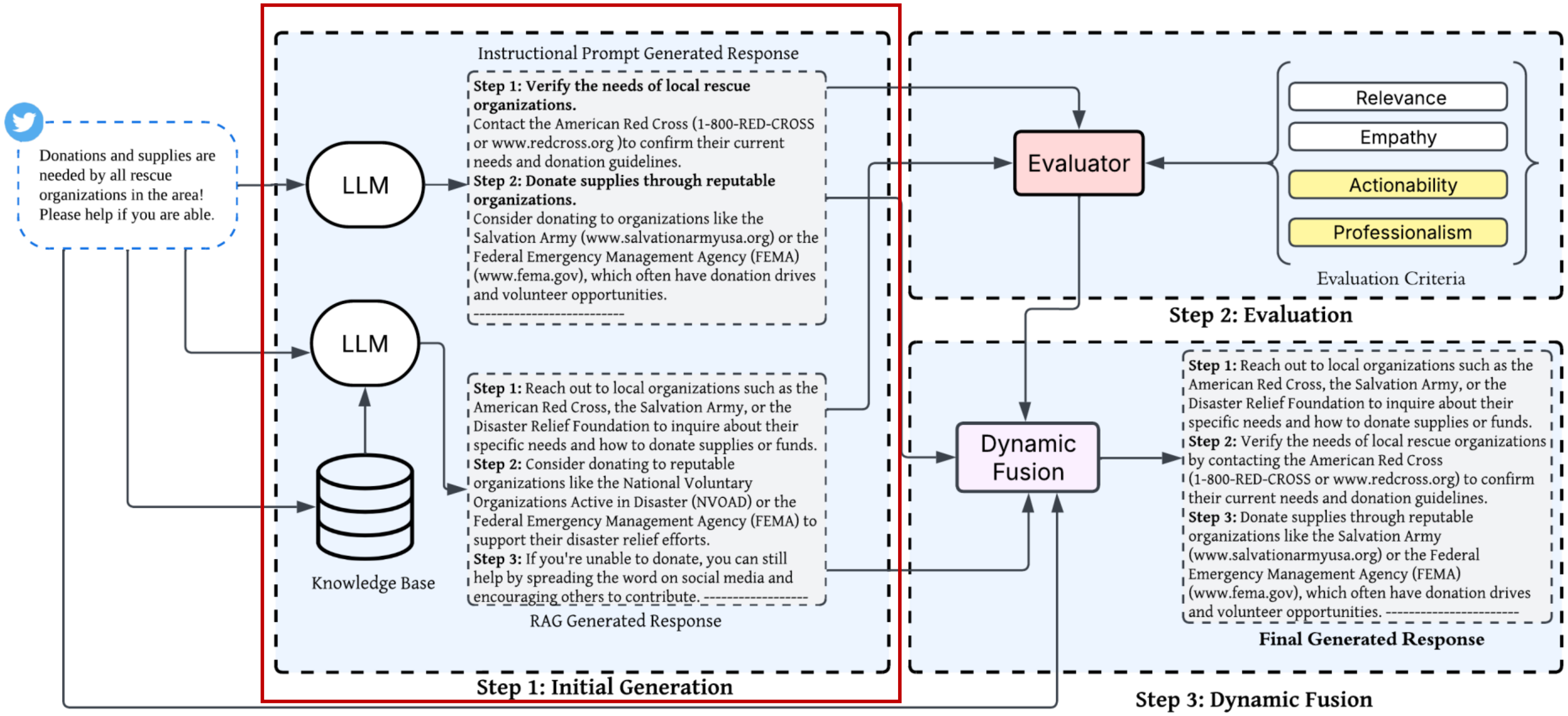
Dynamic Fusion (Our Contribution)



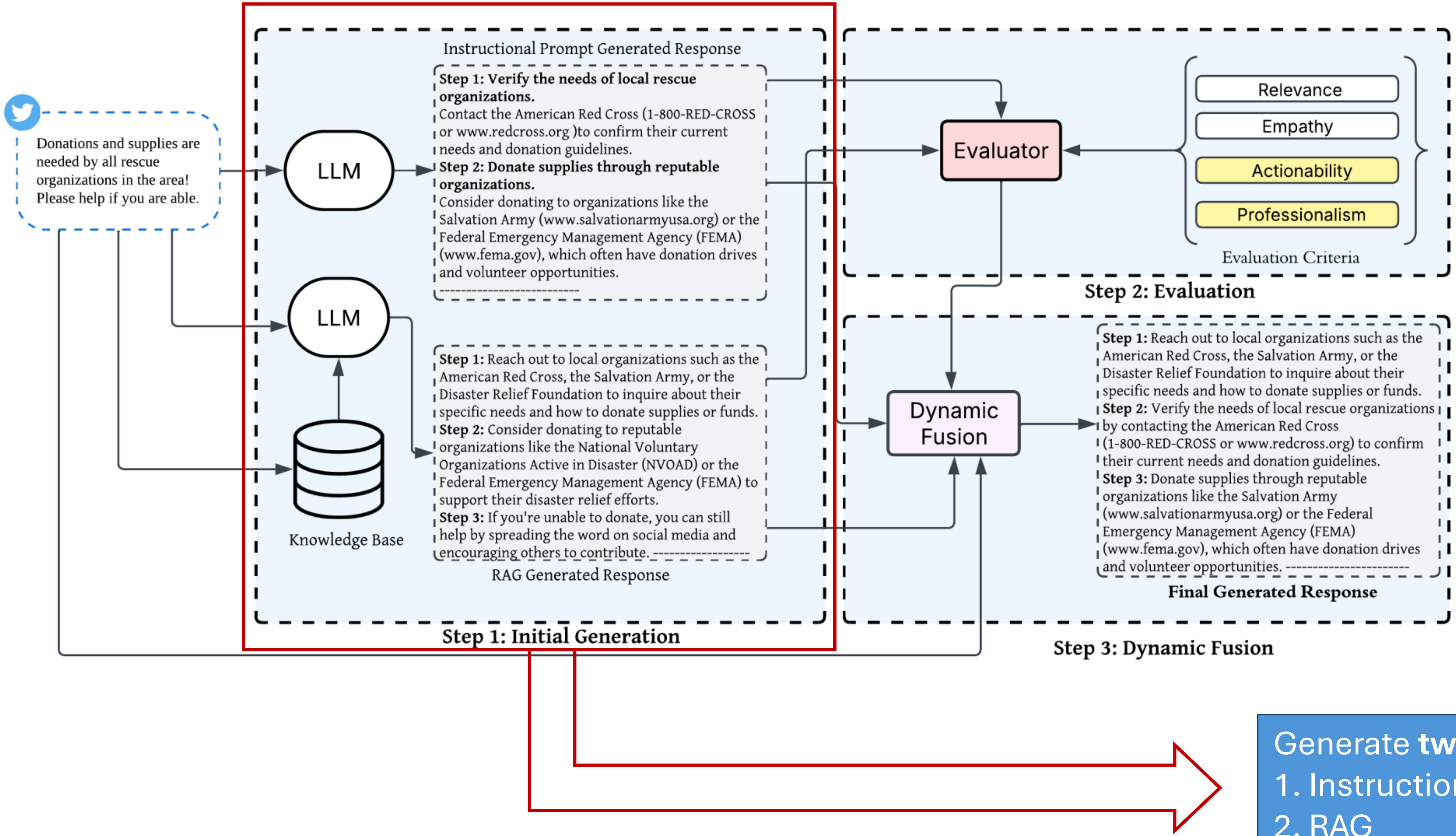
Dynamic Fusion (Our Contribution)



Dynamic Fusion (Our Contribution)

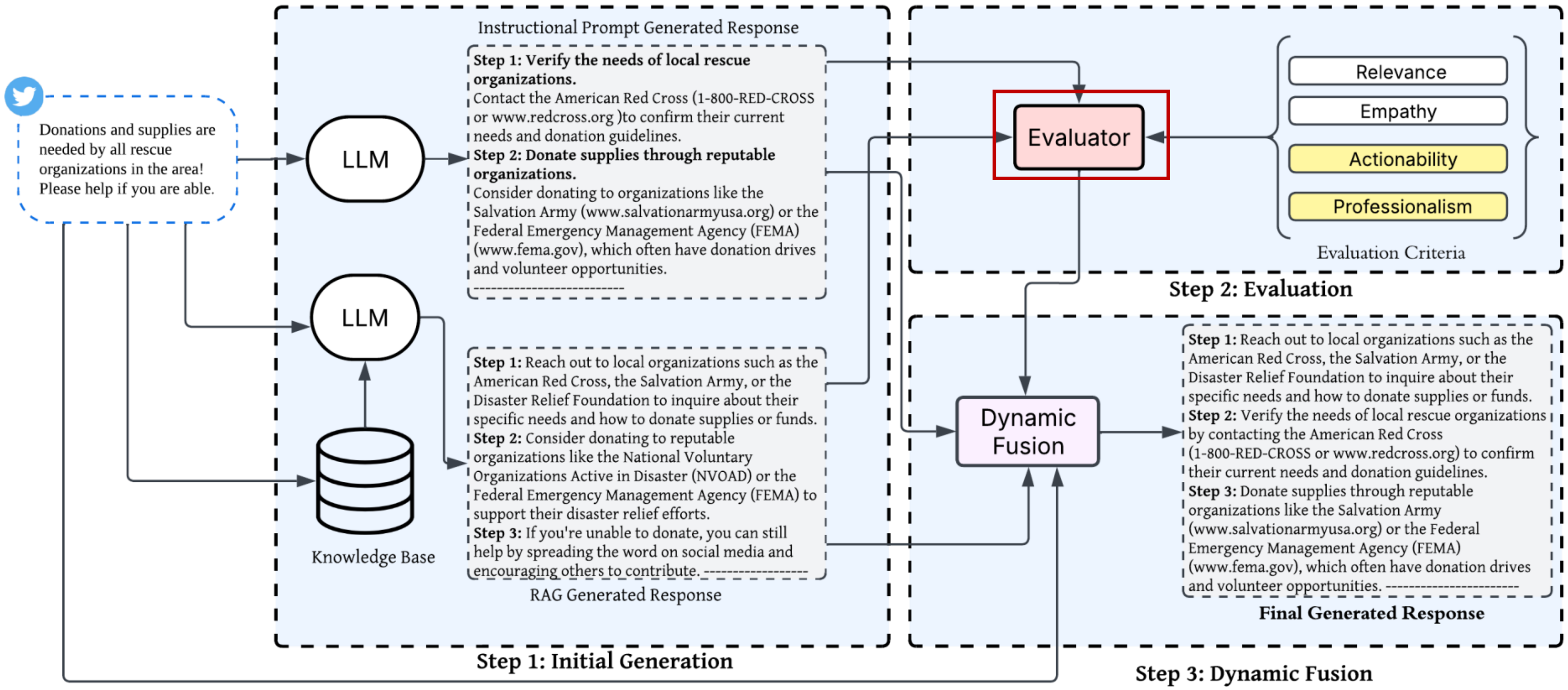


Dynamic Fusion (Our Contribution)

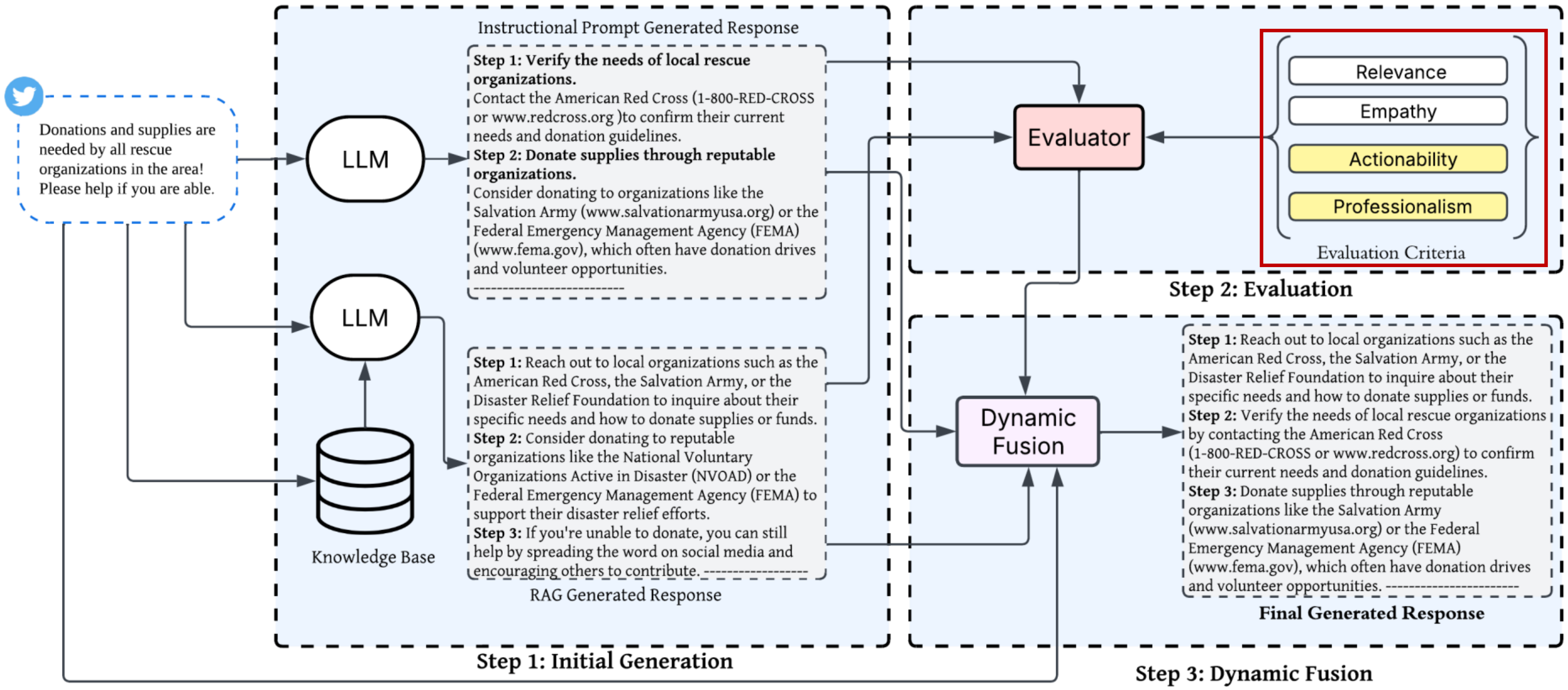


Generate **two** responses:
1. Instruction Prompt
2. RAG

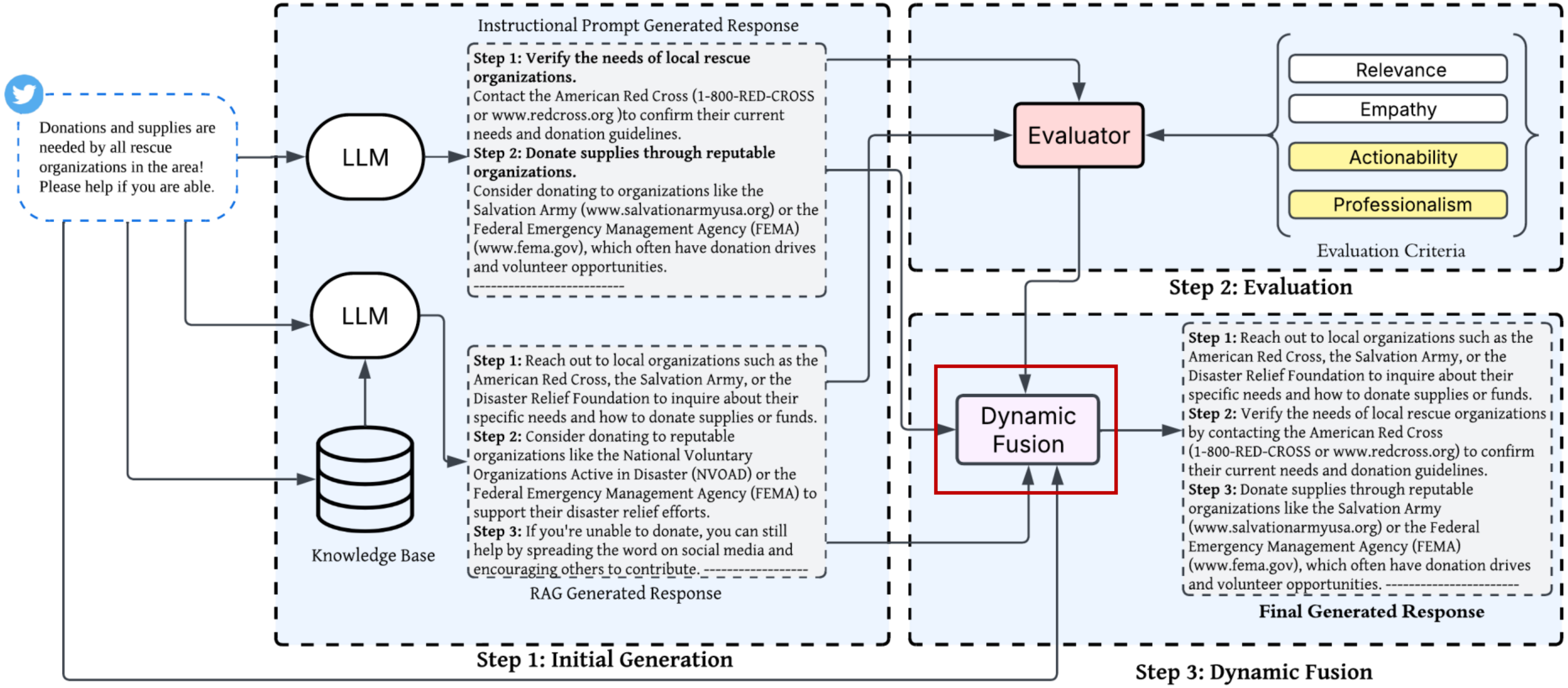
Dynamic Fusion (Our Contribution)



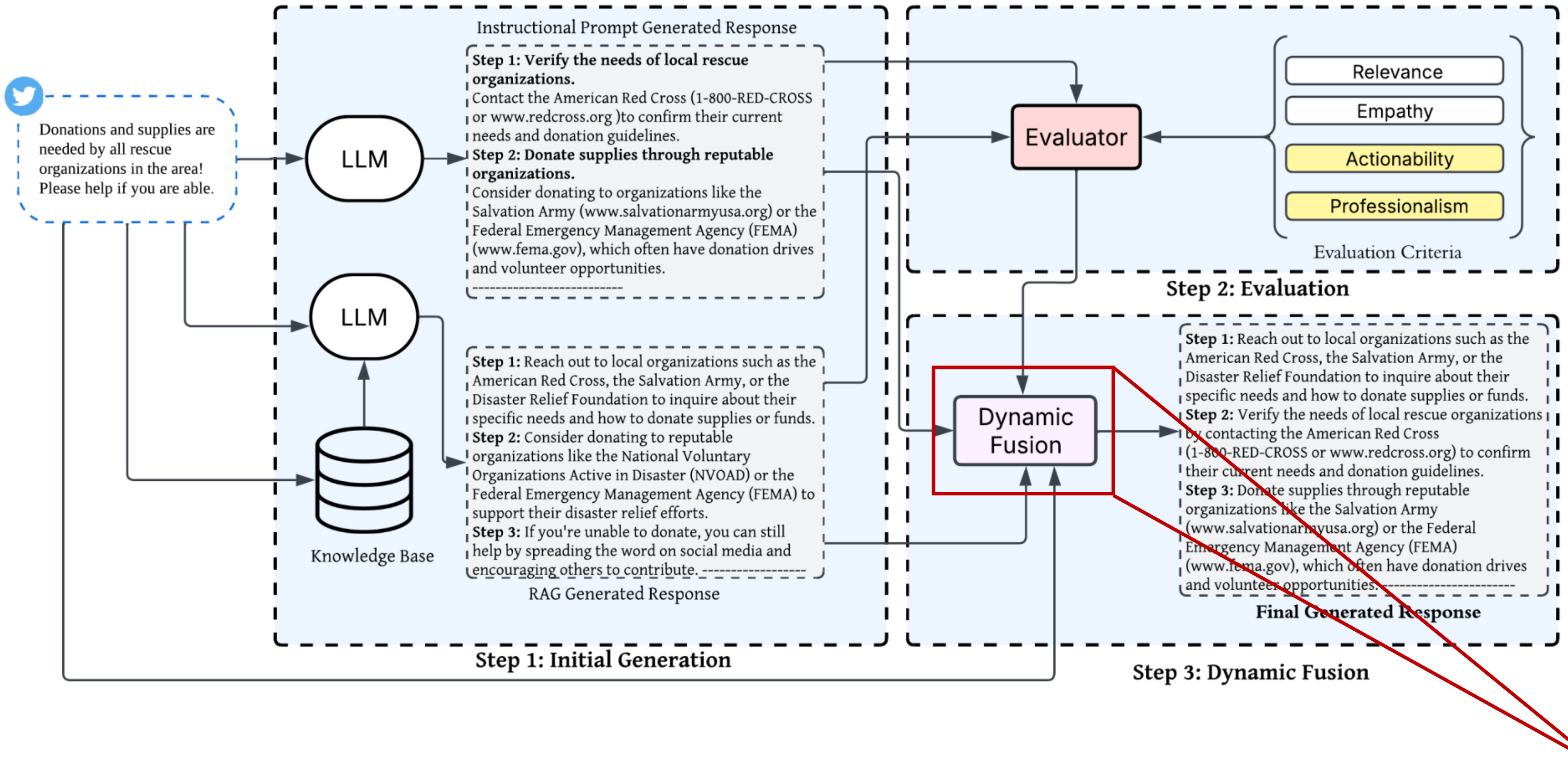
Dynamic Fusion (Our Contribution)



Dynamic Fusion (Our Contribution)

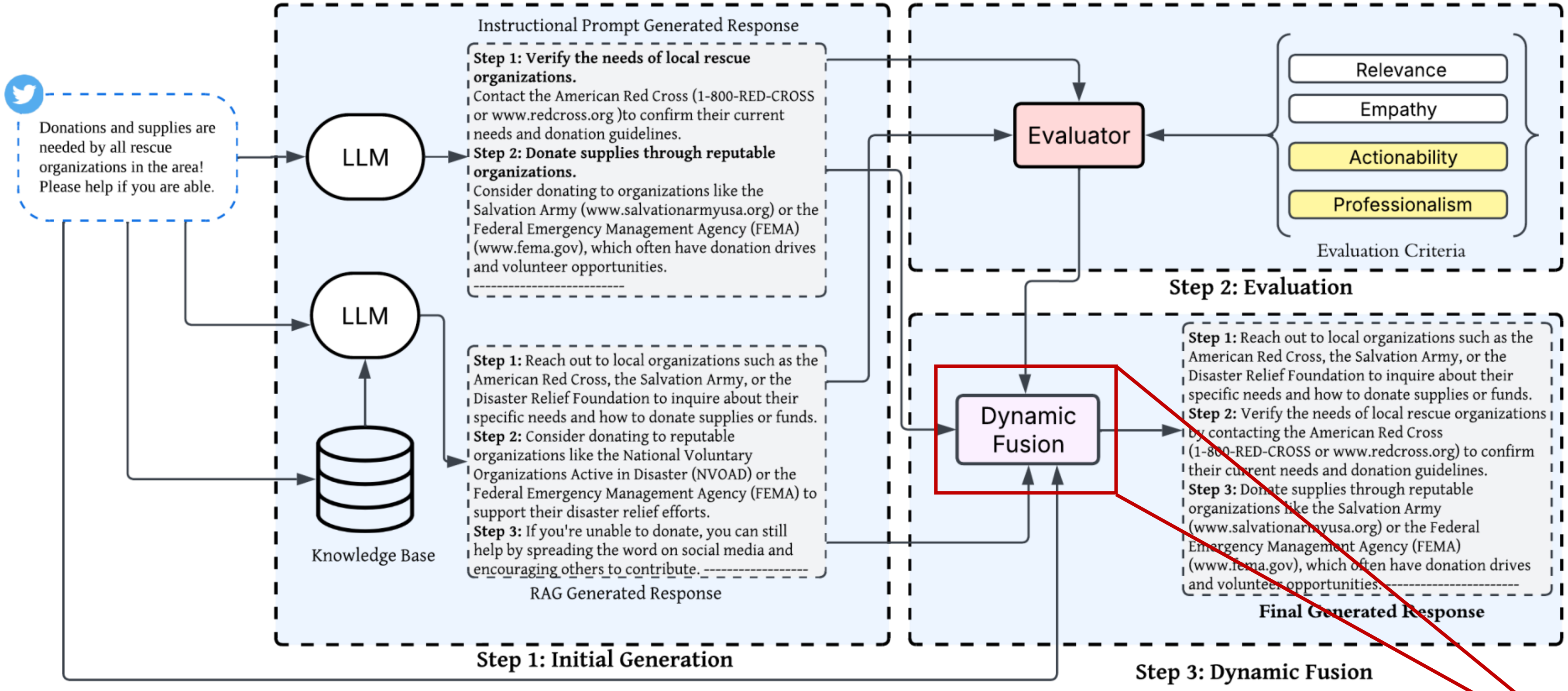


Dynamic Fusion (Our Contribution)



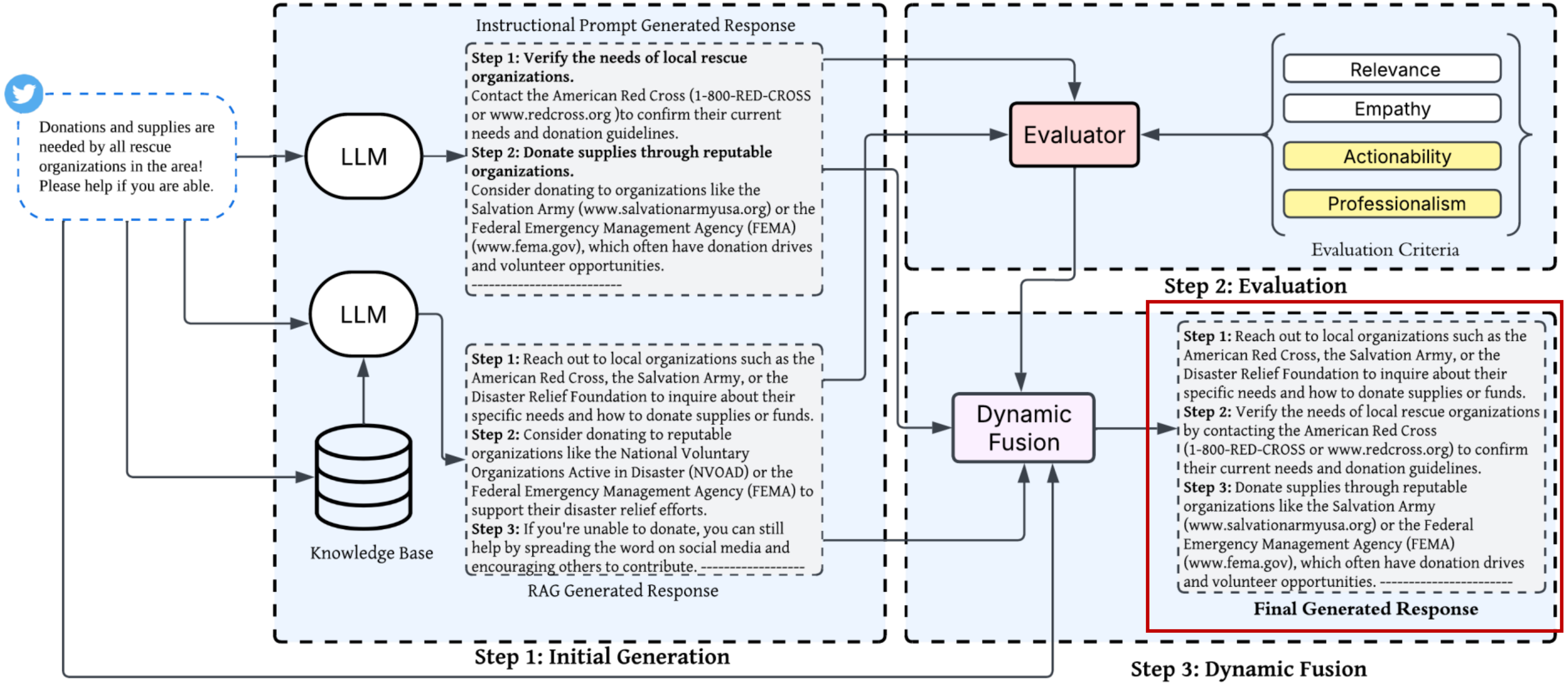
If one response scores highest across all → **Select it**

Dynamic Fusion (Our Contribution)



If mixed performance → **Fuse** the best-scoring parts from each

Dynamic Fusion (Our Contribution)



Results

How Well Do the Methods Perform?
We evaluate responses generated by:

- **Instruction Prompt**
- **Retrieval-Augmented Generation (RAG)**
- **Dynamic Fusion (Proposed)**

Across three LLMs:

- LLaMA 3.1 8B Instruct
- Mistral 8B Instruct
- Qwen2.5 7B Instruct

Results

How Well Do the Methods Perform?
We evaluate responses generated by:

- **Instruction Prompt**
- **Retrieval-Augmented Generation (RAG)**
- **Dynamic Fusion (Proposed)**

Across three LLMs:

- LLaMA 3.1 8B Instruct
- Mistral 8B Instruct
- Qwen2.5 7B Instruct

Model	Method	P	A	E	R	O (Overall)
LLaMA	Instruction Prompt	0.74	0.52	0.06	0.43	0.55
	RAG	0.96	0.63	**0.22**	0.40	0.70
	Dynamic Fusion	**0.92**	**0.97**	0.04	**0.46**	**0.81**
Mistral	Instruction Prompt	0.87	0.98	0.03	0.41	0.78
	RAG	0.87	0.97	0.13	0.42	0.79
	Dynamic Fusion	**0.97**	**0.99**	0.06	**0.50**	**0.84**
Qwen	Instruction Prompt	0.98	0.99	0.03	0.45	0.84
	RAG	0.79	0.86	**0.14**	0.47	0.72
	Dynamic Fusion	**0.98**	**0.99**	0.08	**0.51**	**0.85**

P = Professionalism; **A** = Actionability; **E** = Empathy; **R** = Relevance; **O** = Overall Quality (weighted).

Results

How Well Do the Methods Perform?
We evaluate responses generated by:

- **Instruction Prompt**
- **Retrieval-Augmented Generation (RAG)**
- **Dynamic Fusion (Proposed)**

Across three LLMs:

- LLaMA 3.1 8B Instruct
- Mistral 8B Instruct
- Qwen2.5 7B Instruct

Model	Method	P	A	E	R	O (Overall)
LLaMA	Instruction Prompt	0.74	0.52	0.06	0.43	0.55
	RAG	0.96	0.63	**0.22**	0.40	0.70
	Dynamic Fusion	**0.92**	**0.97**	0.04	**0.46**	**0.81**
Mistral	Instruction Prompt	0.87	0.98	0.03	0.41	0.78
	RAG	0.87	0.97	**0.13**	0.42	0.79
	Dynamic Fusion	**0.97**	**0.99**	0.06	**0.50**	**0.84**
Qwen	Instruction Prompt	0.98	0.99	0.03	0.45	0.84
	RAG	0.79	0.86	**0.14**	0.47	0.72
	Dynamic Fusion	**0.98**	**0.99**	0.08	**0.51**	**0.85**

P = Professionalism; **A** = Actionability; **E** = Empathy; **R** = Relevance; **O** = Overall Quality (weighted).

Results

How Well Do the Methods Perform?
We evaluate responses generated by:

- **Instruction Prompt**
- **Retrieval-Augmented Generation (RAG)**
- **Dynamic Fusion (Proposed)**

Across three LLMs:

- LLaMA 3.1 8B Instruct
- Mistral 8B Instruct
- Qwen2.5 7B Instruct

Model	Method	P	A	E	R	O (Overall)
LLaMA	Instruction Prompt	0.74	0.52	0.06	0.43	0.55
	RAG	0.96	0.63	**0.22**	0.40	0.70
	Dynamic Fusion	**0.92**	**0.97**	0.04	**0.46**	**0.81**
Mistral	Instruction Prompt	0.87	0.98	0.03	0.41	0.78
	RAG	0.87	0.97	**0.13**	0.42	0.79
	Dynamic Fusion	**0.97**	**0.99**	0.06	**0.50**	**0.84**
Qwen	Instruction Prompt	0.98	0.99	0.03	0.45	0.84
	RAG	0.79	0.86	**0.14**	0.47	0.72
	Dynamic Fusion	**0.98**	**0.99**	0.08	**0.51**	**0.85**

P = Professionalism; **A** = Actionability; **E** = Empathy; **R** = Relevance; **O** = Overall Quality (weighted).

Results

How Well Do the Methods Perform?
We evaluate responses generated by:

- **Instruction Prompt**
- **Retrieval-Augmented Generation (RAG)**
- **Dynamic Fusion (Proposed)**

Across three LLMs:

- LLaMA 3.1 8B Instruct
- Mistral 8B Instruct
- Qwen2.5 7B Instruct

Model	Method	P	A	E	R	O (Overall)
LLaMA	Instruction Prompt	0.74	0.52	0.06	0.43	0.55
	RAG	0.96	0.63	**0.22**	0.40	0.70
	Dynamic Fusion	**0.92**	**0.97**	0.04	**0.46**	**0.81**
Mistral	Instruction Prompt	0.87	0.98	0.03	0.41	0.78
	RAG	0.87	0.97	0.13	0.42	0.79
	Dynamic Fusion	**0.97**	**0.99**	0.06	**0.50**	**0.84**
Qwen	Instruction Prompt	0.98	0.99	0.03	0.45	0.84
	RAG	0.79	0.86	**0.14**	0.47	0.72
	Dynamic Fusion	**0.98**	**0.99**	0.08	**0.51**	**0.85**

P = Professionalism; **A** = Actionability; **E** = Empathy; **R** = Relevance; **O** = Overall Quality (weighted).

Results

How Well Do the Methods Perform?
We evaluate responses generated by:

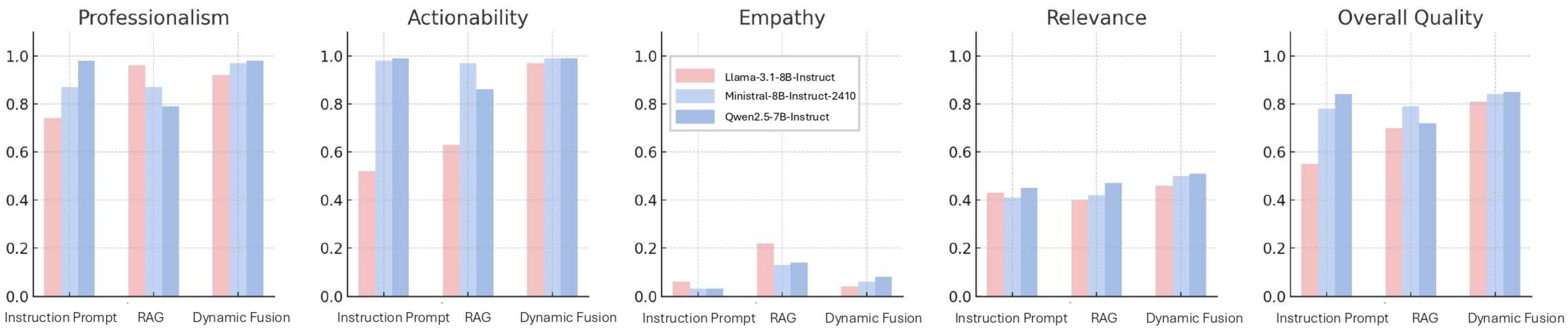
- **Instruction Prompt**
- **Retrieval-Augmented Generation (RAG)**
- **Dynamic Fusion (Proposed)**

Across three LLMs:

- LLaMA 3.1 8B Instruct
- Mistral 8B Instruct
- Qwen2.5 7B Instruct

Model	Method	P	A	E	R	O (Overall)
LLaMA	Instruction Prompt	0.74	0.52	0.06	0.43	0.55
	RAG	0.96	0.63	**0.22**	0.40	0.70
	Dynamic Fusion	**0.92**	**0.97**	0.04	**0.46**	**0.81**
Mistral	Instruction Prompt	0.87	0.98	0.03	0.41	0.78
	RAG	0.87	0.97	0.13	0.42	0.79
	Dynamic Fusion	**0.97**	**0.99**	0.06	**0.50**	**0.84**
Qwen	Instruction Prompt	0.98	0.99	0.03	0.45	0.84
	RAG	0.79	0.86	**0.14**	0.47	0.72
	Dynamic Fusion	**0.98**	**0.99**	0.08	**0.51**	**0.85**

P = Professionalism; **A** = Actionability; **E** = Empathy; **R** = Relevance; **O** = Overall Quality (weighted).



Conclusions

Key Findings

- **Large Language Models (LLMs)** show strong potential for assisting in real-time crisis communication on social media.
- **Instruction Prompt** and **RAG** methods each have strengths:
    - Instruction Prompts offer structure response.
    - RAG improves factual grounding and reduces hallucinations.
- Our proposed **Dynamic Fusion** method consistently delivers the **highest overall response quality**, combining:
    - Professional tone
    - Clear, actionable guidance
    - Grounded information
    - Balanced relevance

Conclusions

Key Findings

- **Large Language Models (LLMs)** show strong potential for assisting in real-time crisis communication on social media.
- **Instruction Prompt** and **RAG** methods each have strengths:
    - Instruction Prompts offer structure response.
    - RAG improves factual grounding and reduces hallucinations.
- Our proposed **Dynamic Fusion** method consistently delivers the **highest overall response quality**, combining:
    - Professional tone
    - Clear, actionable guidance
    - Grounded information
    - Balanced relevance

Dynamic Fusion offers a promising path to building ***scalable, trustworthy AI responders*** *that can support affected individuals during crises.*

Limitations & Future Work

Current Limitations

- **Single Crisis Scenario**
  - Tested only on **Hurricane Irma** tweets
  - May not generalize across **other disaster types** (e.g., earthquakes, pandemics)
- **Fusion Agent Simplicity**
  - Fusion currently relies on **basic score comparison and selection**
  - Might overlook deeper reasoning or nuanced combinations
- **Empathy Gap**
  - Dynamic Fusion improved professionalism & actionability
  - But **still lags in empathetic tone** compared to human responders

Limitations & Future Work

Current Limitations

- **Single Crisis Scenario**
  - Tested only on **Hurricane Irma** tweets
  - May not generalize across **other disaster types** (e.g., earthquakes, pandemics)
- **Fusion Agent Simplicity**
  - Fusion currently relies on **basic score comparison and selection**
  - Might overlook deeper reasoning or nuanced combinations
- **Empathy Gap**
  - Dynamic Fusion improved professionalism & actionability
  - But **still lags in empathetic tone** compared to human responders

Future Work

- **Expand to multiple crisis types**
  - Apply to floods, wildfires, and health emergencies
  - Test multilingual and multicultural variations
- **Improve Fusion Strategy**
  - Integrate **Chain-of-Thought (CoT)** reasoning to guide how components are fused
  - Explore **adaptive prompting or ensemble decision-making**
- **Real-Time Deployment Potential**
  - Collaborate with **NGOs or emergency response teams**
  - Enable LLMs to act as **first-line information assistants**

Thank You!

Contact: AnirbanSahaAnik@my.unt.edu

Acknowledgment
This work was supported by the Institute of Museum and Library Services (IMLS) National Leadership Grants under LG256661-OLS-24 and LG-256666-OLS-24.