

## ABSTRACT

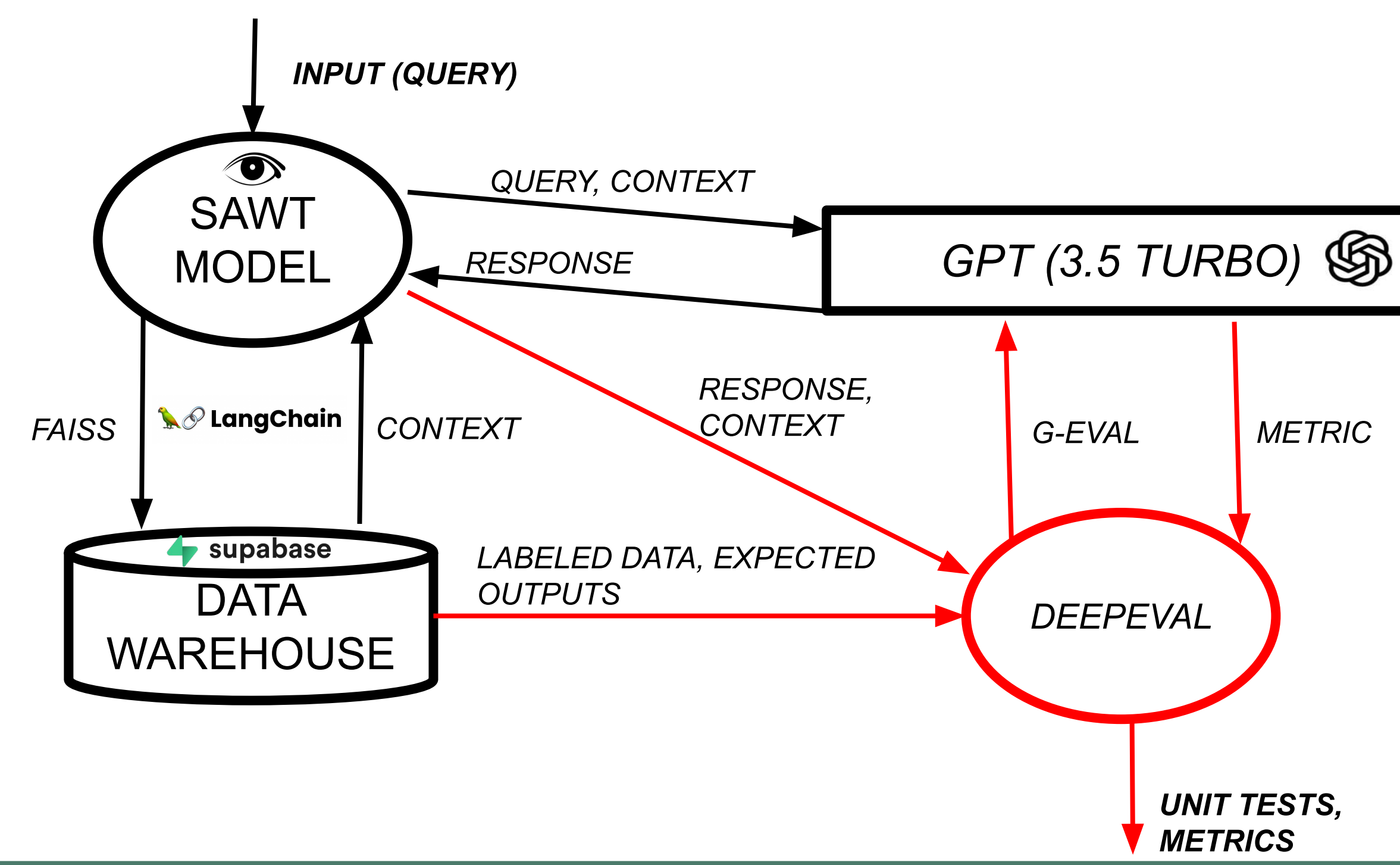
Eye on Surveillance, a New Orleans-based community group, has developed Sawt, an Large Language Model (LLM)-backed tool generating responses to user queries, built on Retrieval Augmented Generation (RAG) context from New Orleans City Council information to enhance transparency and accessibility of government information.

Given concerns regarding AI systems' accuracy, bias, and equity, it's crucial for public-facing systems like Sawt to undergo rigorous evaluation and community feedback to ensure trustworthiness. Key considerations include response accuracy, fair representation of diverse perspectives, and the impact of various LLM design choices (e.g., model selection, prompts, source documents). This project aims to address two primary research questions:

**RQ1: How helpful, accurate, and balanced are AI-generated responses in Sawt?**

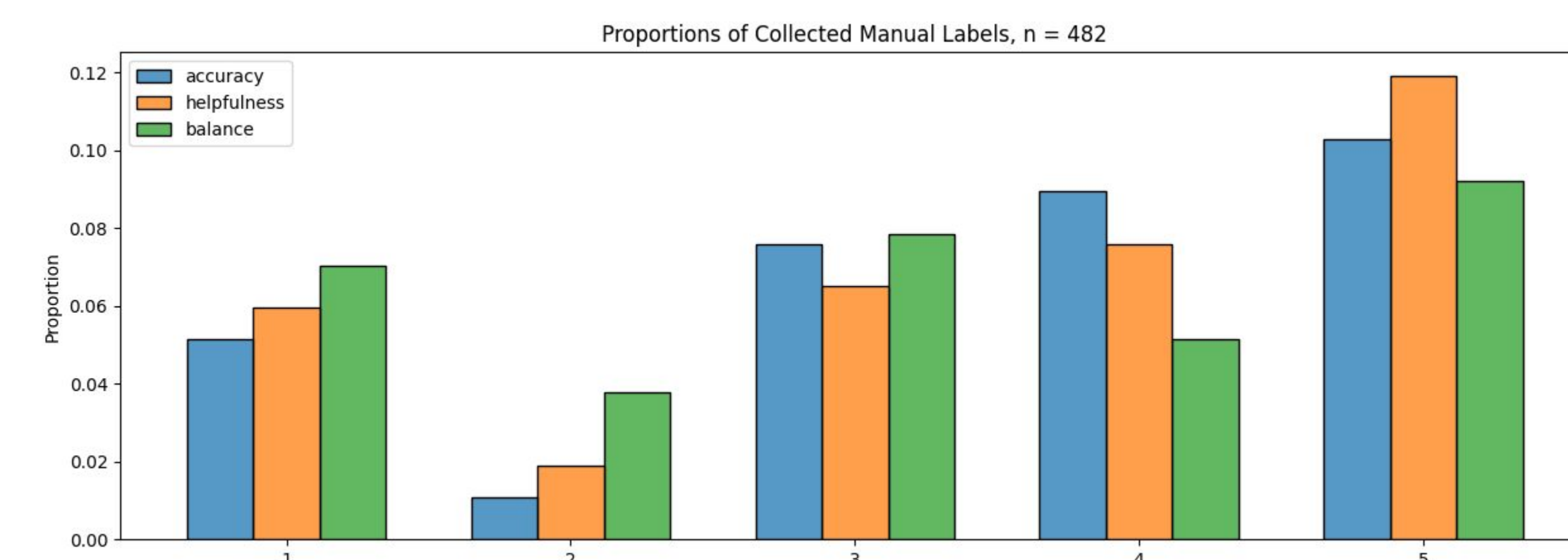
**RQ2: What is the effectiveness of automated evaluation tools in assessing RAG system quality?**

## ARCHITECTURE



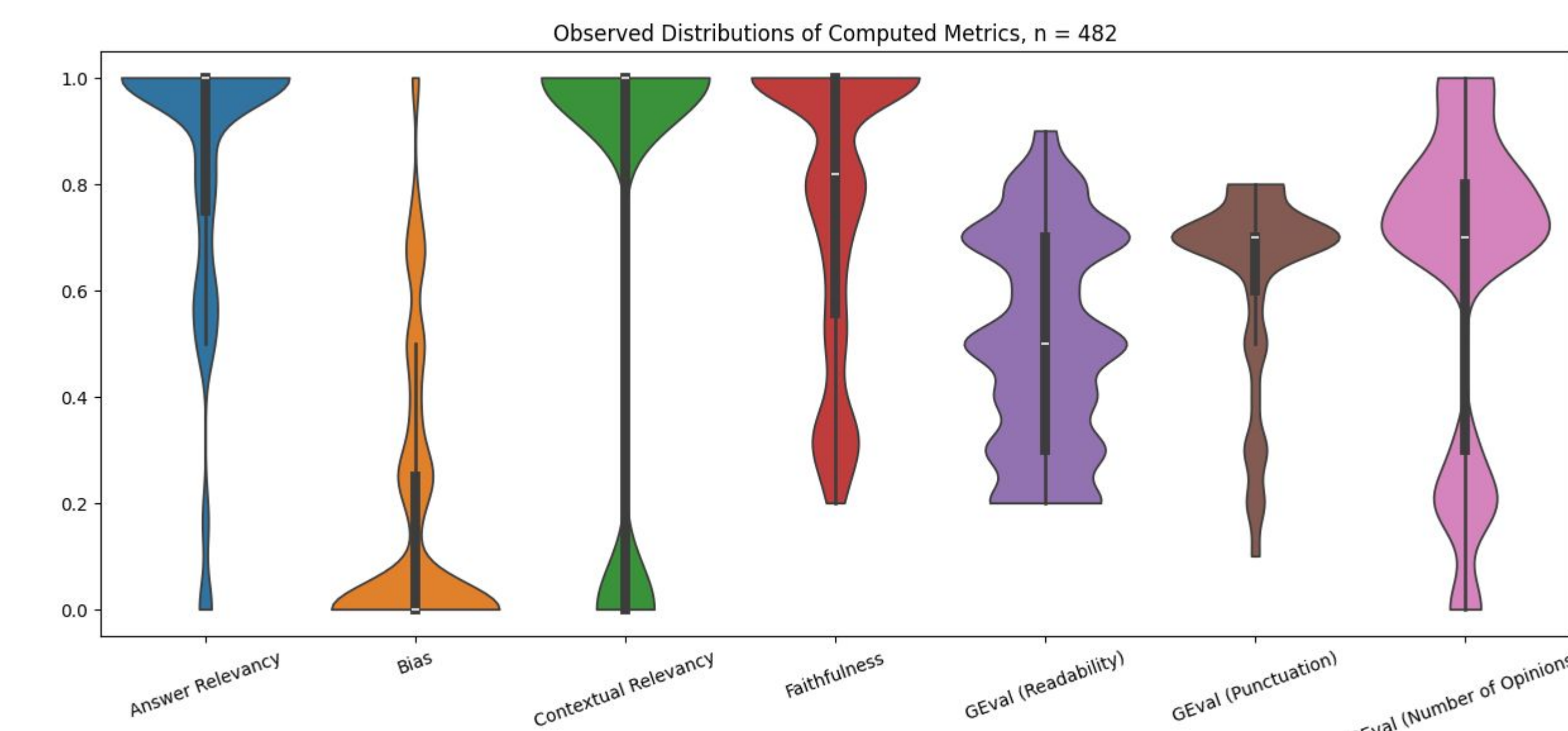
## USER FEEDBACK DATA

- Collected across two community engagement sessions with Eye on Surveillance
- Consisted of local activists, citizens, and organizers
- Rated cached queries with three distinct responses with a varied number of vectorized context documents returned for response generation
- Evaluated with respect to "Helpfulness", "Accuracy", and "Balance" of response on a discrete scale 1-5



## AUTOMATED MODEL EVALUATION (DeepEval)

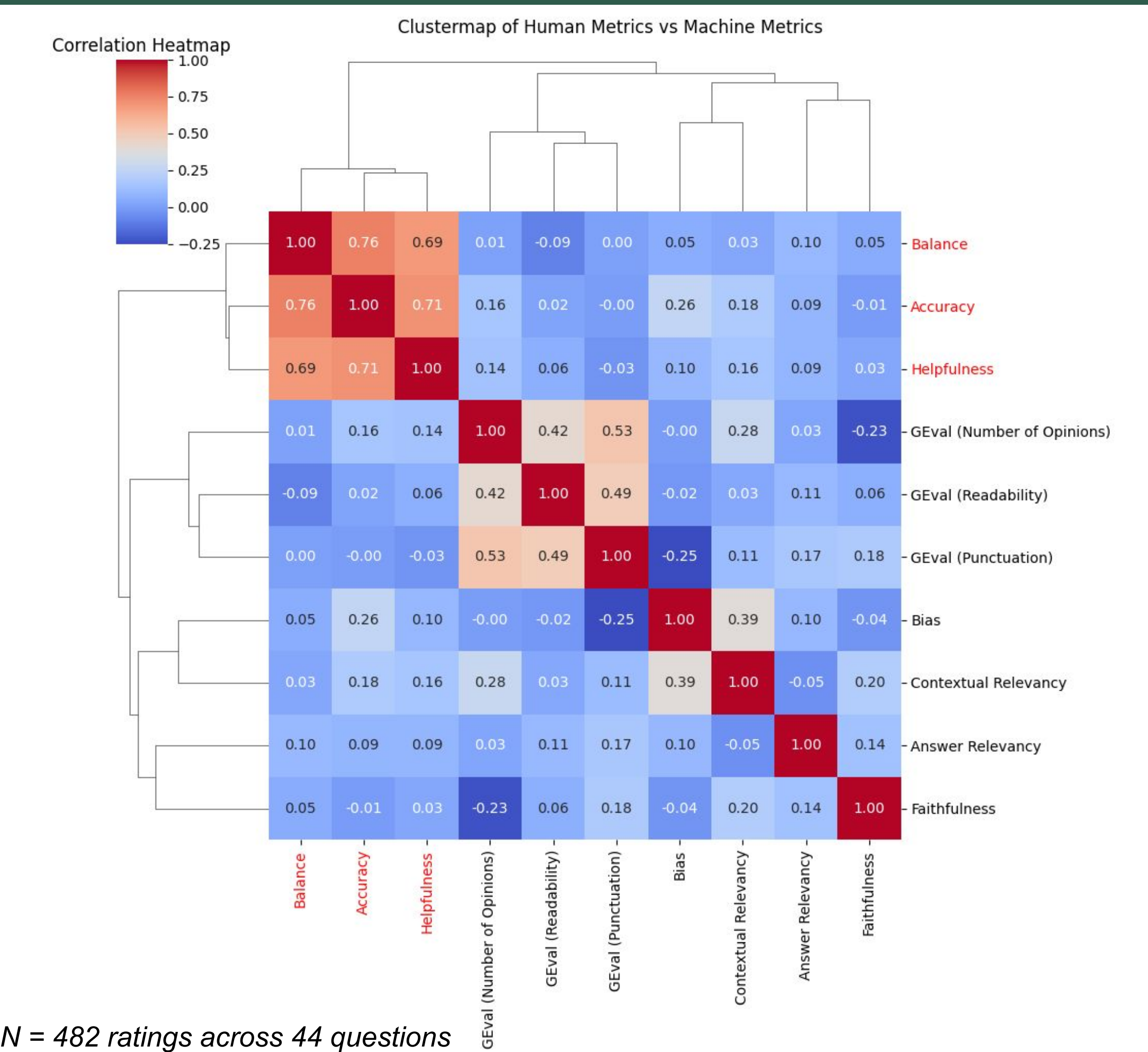
- 7 Metrics designed to evaluate the quality of the RAG pipeline
- Answer Relevancy, Faithfulness, Bias, Contextual Relevancy, Readability, Punctuation, Number of Opinions
- PyTest Unit tests to evaluate model responses
- Hyperparameter comparison to deduce best model configurations
- Incorporated human User-Feedback data to augment the evaluation system
- Provide a frontend interface for monitoring output quality over response data and time



## DATA SOURCES

- New Orleans City Council Meeting Transcripts, Minutes, and Agendas
- Criminal Justice Meeting Transcripts, Minutes, and Agendas
- Public Comment Submissions
- Local News Sources

## METRIC SIMILARITIES



### Notable Positive Correlations

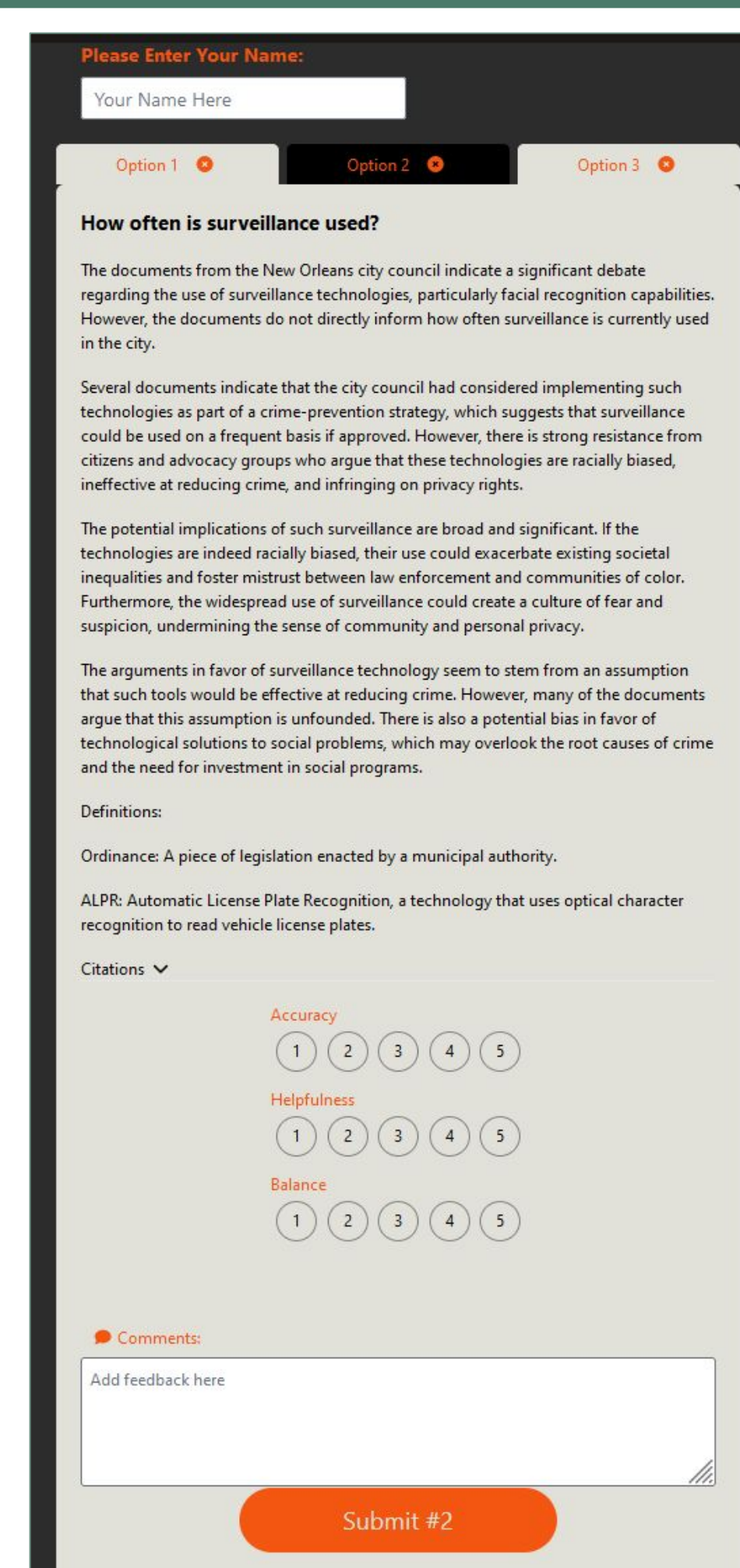
- **Labeled Accuracy & Computed Bias + Contextual Relevancy**
- **Labeled Helpfulness and Computed # Of Opinions (GEVAL)**
- **All GEVAL custom metrics**

## FUTURE RESEARCH

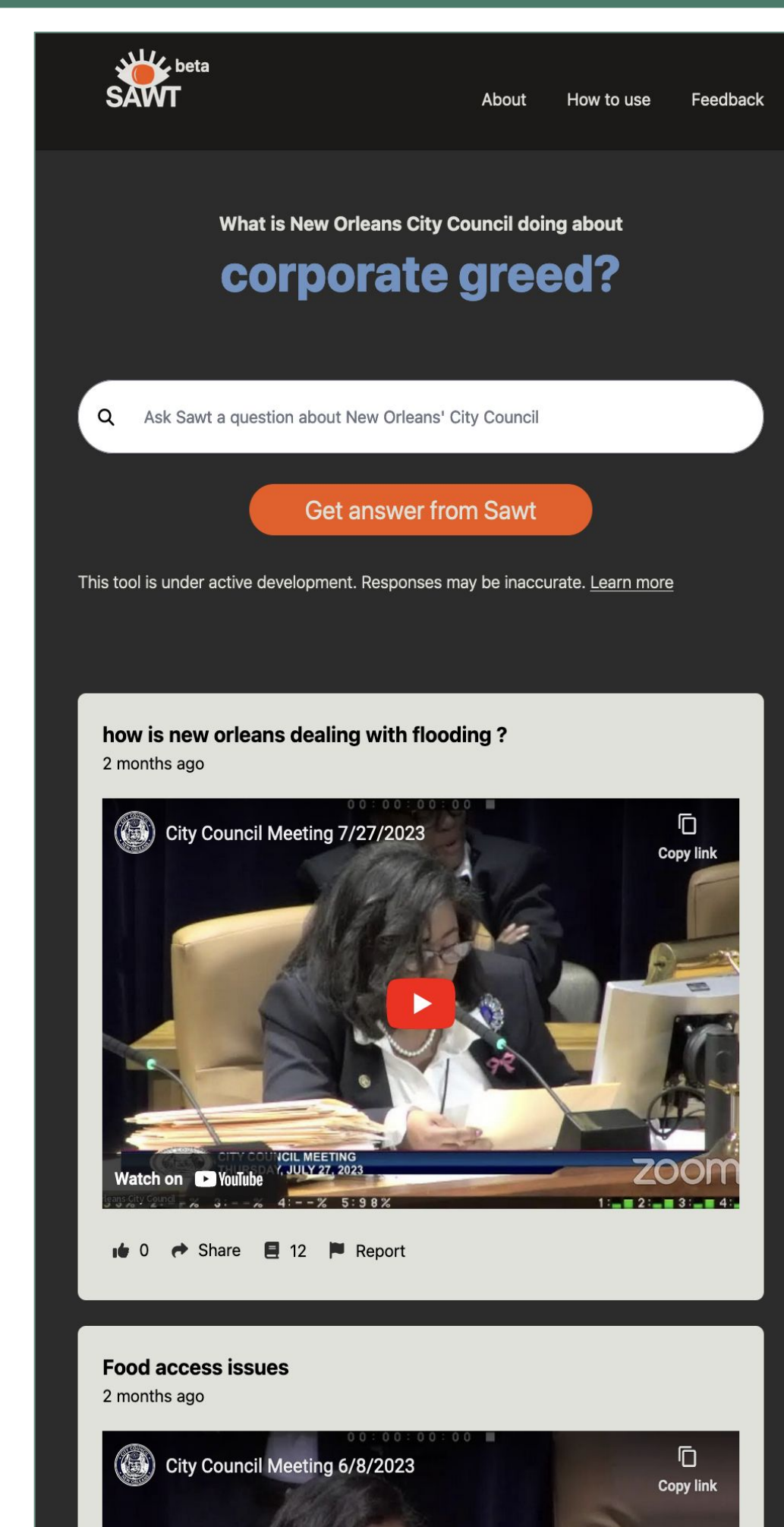
- Hyperparameter Tuning
- Continuous Integration/Deployment
- Additional Community Data Collection
- Multiple Query Formats
- Refining Metric Uncertainty and Calculation Process

### THANKS TO:

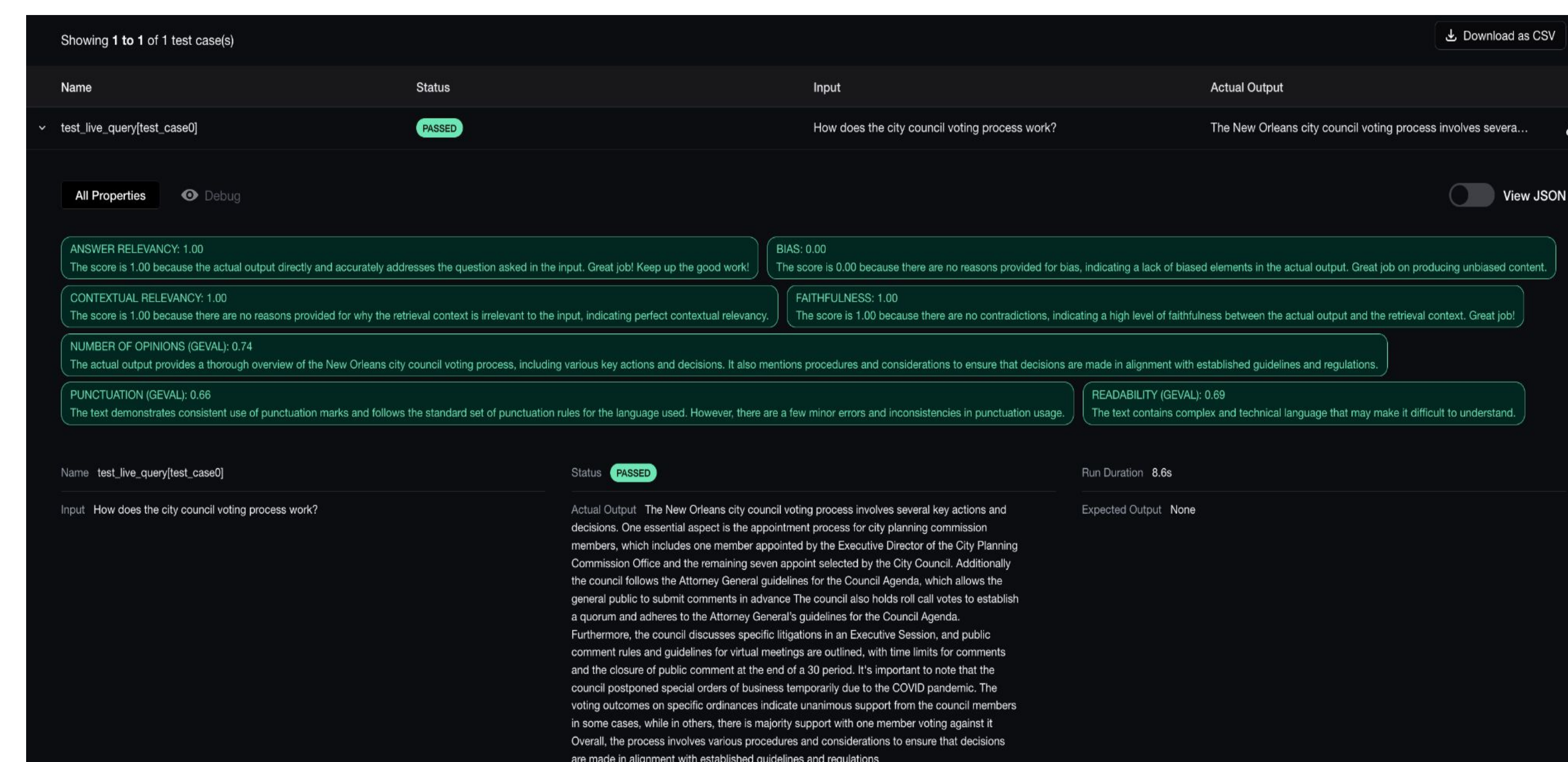
- Eye on Surveillance New Orleans
- Tulane University Computer Science & Center for Community Engaged Artificial Intelligence



User Feedback Data Collection Tool



Sawt Interface (EOS)



Demo Query Response Metrics in DeepEval Frontend

<https://sawt.us>



<https://eyeonsurveillance.org>



<https://sse.tulane.edu/cs/ceai>

