

CLARITY - Sensemaking

—

Intern Team 2025



Agenda

Project Inspiration

Clarity's Connection to *Sensemaking Reimagined*

Data Collection

Retrieval-Augmented Generation

Agentic Workflow and Agents

Demonstration

Future Work & Conclusion

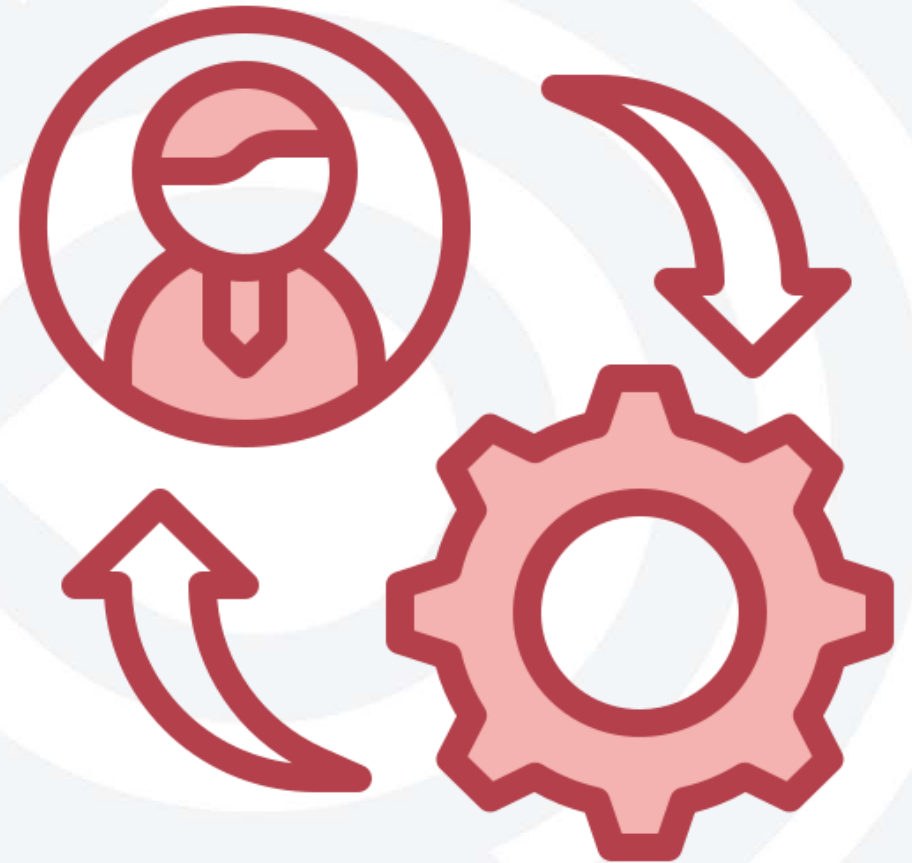
Workflow

❑ Agile Methodology

- Two Week sprints
- Standups
- Sprint planning meetings
- Retrospectives
- Technical exchange meetings between teams

❑ Alternating between roles

- Software Engineer, Prompt Engineer, Database Administrator, Data Scraper, AI Engineer



Traditional Analytics

- Grace tracks China's military, technological, economic, and political activities
- Sequential intelligence cycle
- Critical links often missed
- Pace of reporting causes reactive decision-making

A woman with dark hair in braids, wearing glasses and a small earring, is shown in profile, looking towards the right. She is in front of a large digital display showing various data visualizations, including bar charts and line graphs, with a blue and green color scheme.

Sensemaking REIMAGINED

See How Next-Gen Systems Can Outpace Urgent Threats

Don Polaski, Marissa Beall, and Christopher Castelli

Next Generation Sensemaking

- Grace focuses on insight and providing strategic guidance to system
- Continuous, AI-enabled data fusion across domains
- Real-time pattern detection allows for threat anticipation

Reactive → Proactive

A woman with glasses looking at a digital display with data visualizations.

Sensemaking REIMAGINED

See How Next-Gen Systems Can Outpace Urgent Threats

Don Polaski, Marissa Beall, and Christopher Castelli

From Analyst to Brand Manager

- **Instead of** an analyst → **a brand manager** protecting image
- **Instead of** SIGINT → **OSINT** from social, news, open web
- **Instead of** military activity → **narratives, events, entities**
- **Instead of** situational awareness → **total sentiment awareness**
- **Instead of** adversary movements → **reputational spikes** to act on

Reactive to Proactive: Entertainment Sensemaking

Current

Fragmented monitoring

Delayed awareness

Reactive crisis response

Missed opportunities

Reimagined

Always-on sentiment tracking

Unified OSINT streams

Immediate narrative insights

Timely action

Why Reimagined Sensemaking Matters

1

**Early Threat
Detection**

2

**Opportunity
Amplification**

3

**Real-time
Insights**

4

**Narrative
Control**

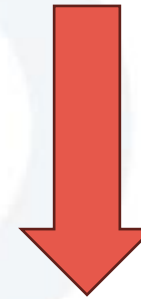
5

**Reputation
Resilience**

OSINT Ingestion
(social, news, metadata)



Event & Entity Linking
(context building)

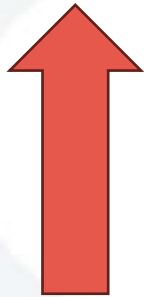


Sentiment Spike Detection
(positive & negative)



Narrative Agent Actions
(mitigation or amplification)

How We Do It



Data Collection



OSINT Sources

- APIs
 - Reddit
 - NewsAPI
- Scraped Data
 - News Sources
 - Fan Pages
 - Music Industry Sites

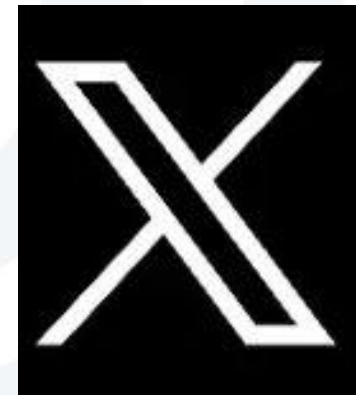


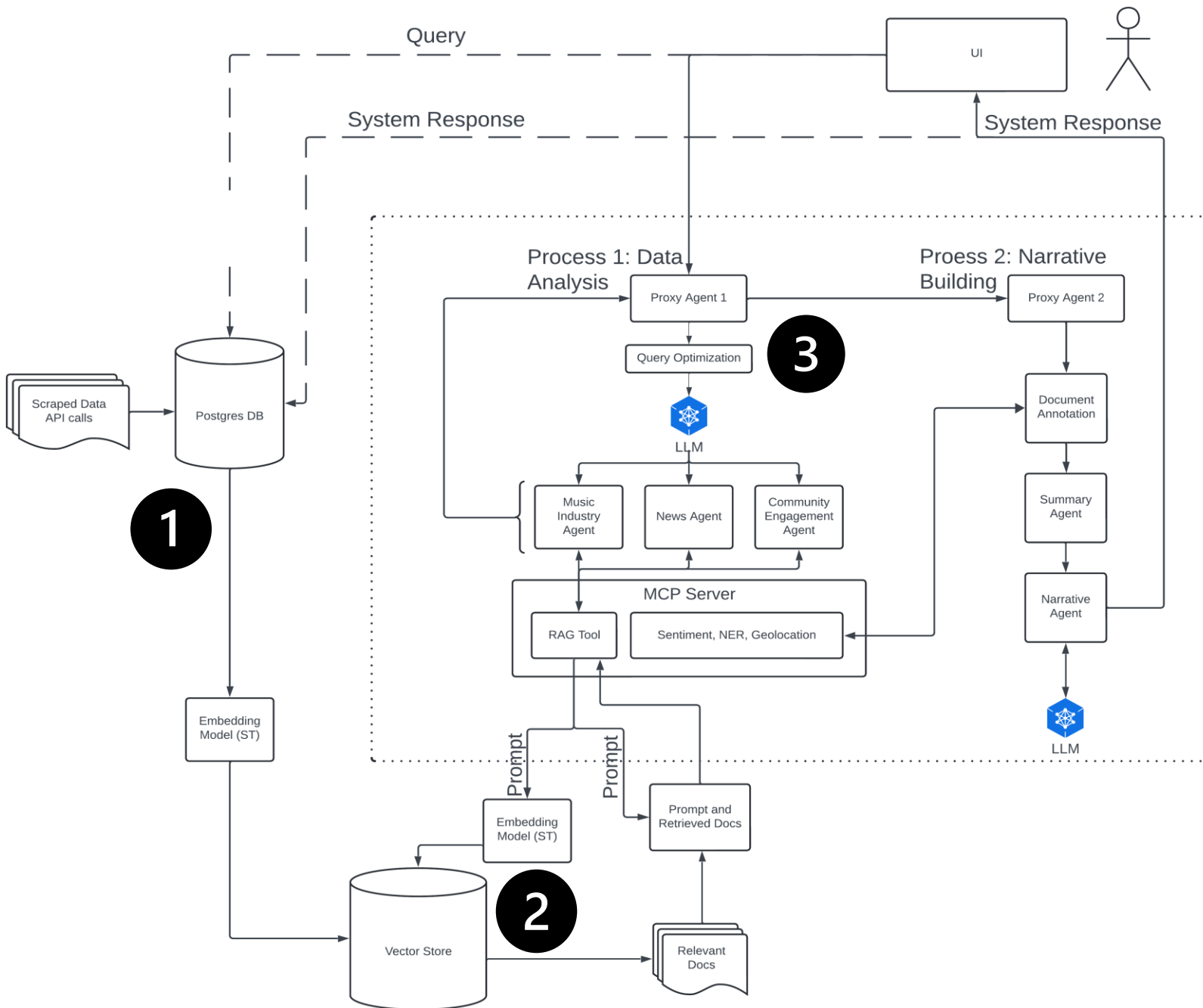
NME

TMZ

Apify-Retrieved Data

- YouTube
 - Descriptions
 - Captions
 - Published date
- Twitter (X)
 - Tweet
 - Hashtag
 - Created date
 - Username
 - Counts: Likes, retweets, replies, quotes





Technologies Used

- PostgreSQL
- Sentence Transformers
- ChromaDB
- Locally hosted Deepseek R1
- LangGraph + LangSmith
- Model Context Protocol (MCP)
- Streamlit
- EC2 instance: g5.2xlarge
- GPU: a10 with 24gb VRAM

Retrieval Augmented Generation (RAG)

LLMs have a problem – hallucinations, cannot answer past cutoffs

```
ubuntu@ip-172-31-46-6:~$ ollama run deepseek-r1:latest  
>>> How was the Stray Kids concert on June 23, 2025?  
I don't have information on specific concerts unless they've been reported in recent events or reviews that  
I could access as of my last training data cutoff (around July 2024). June 23, 2025, is future-dated beyond  
what's  
available to me.
```

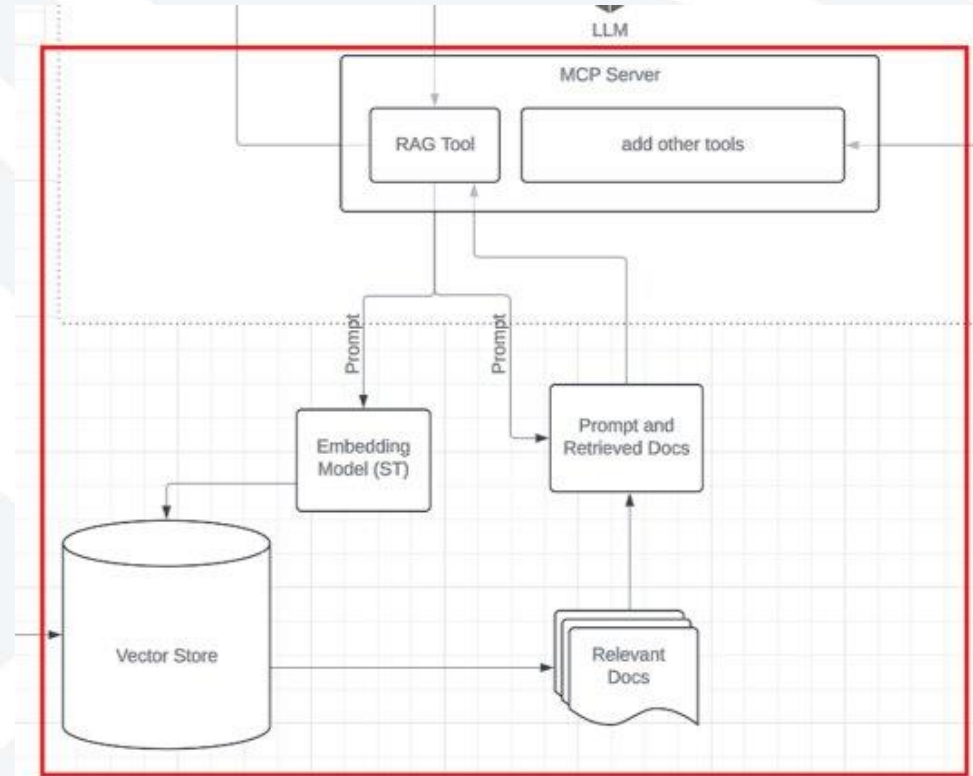
RAG provides external knowledge to the LLM

Retrieval Augmented Generation (RAG)

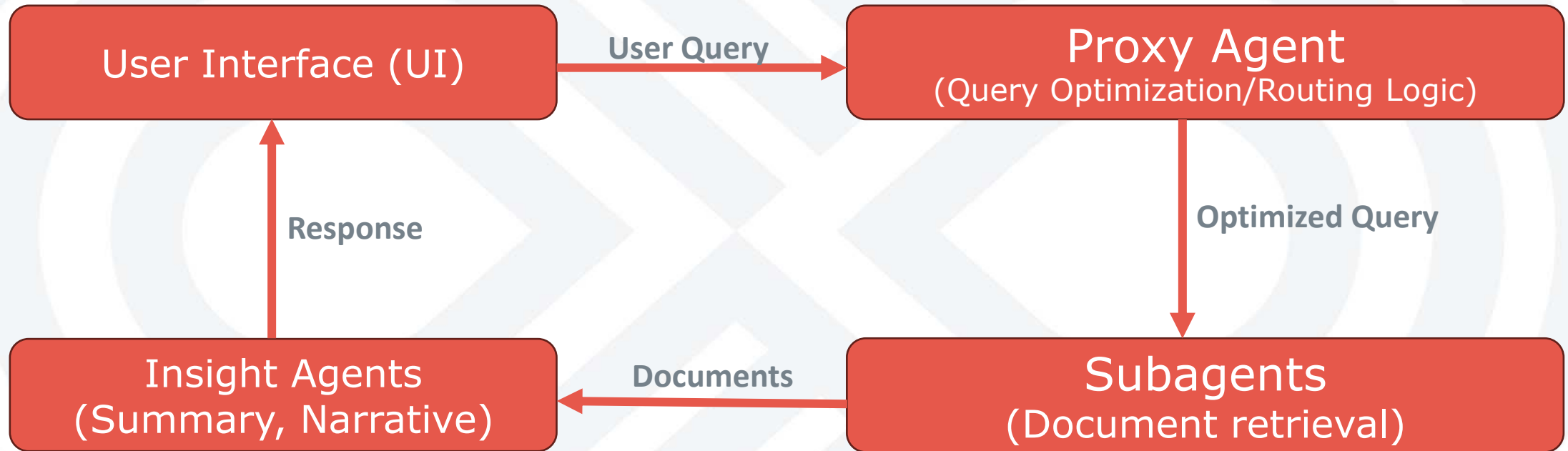
Embeddings are Stored

Retrieval is Initiated

Similarity Search
Retrieves Documents

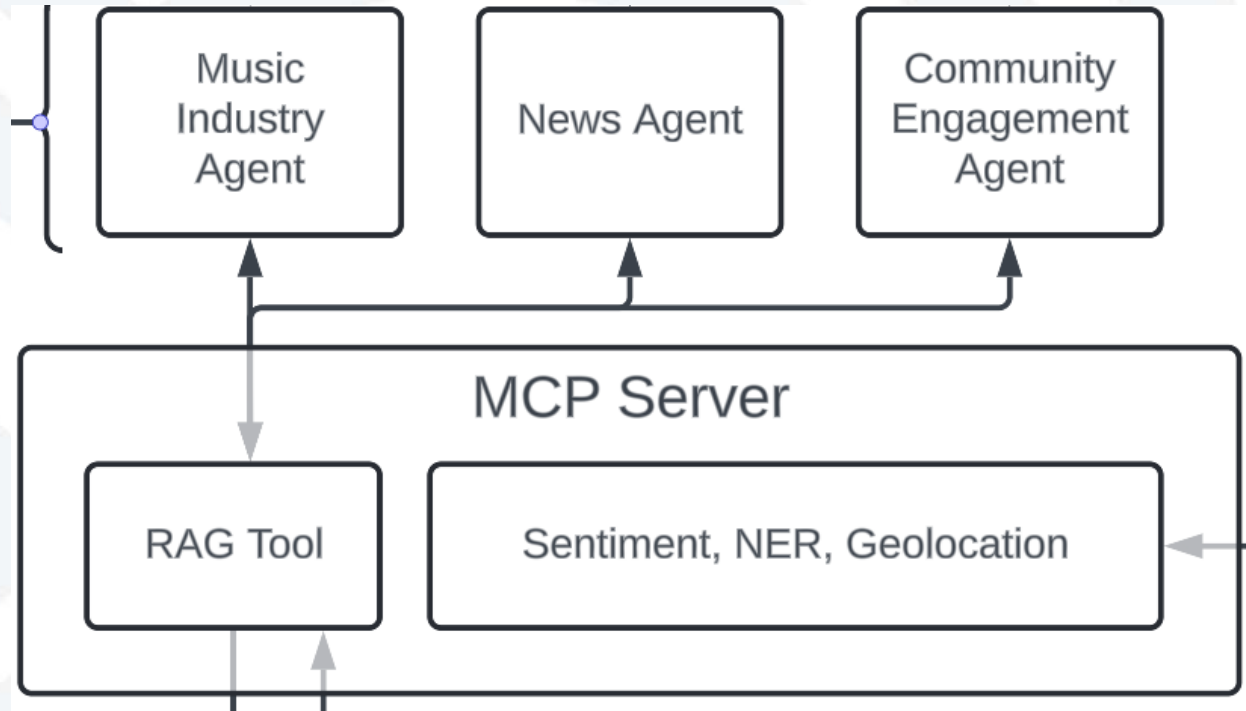


Agentic Workflow



Subagent Logic

1. Send optimized query to RAG tool within MCP server
2. Specific subagents call different RAG embeddings
3. Retrieved documents returned



Summarization Agent



- Selects relevant tools, timeboxing
 - Geolocation
 - NER
 - Sentiment
- Present key takeaways on sentiment, future prediction

```
[summarization_agent] LLM selected tools for community: ['sentiment']  
[{'source': 'twitter_embeddings', 'document': 'HypeMY Stray Kids Bang Chan Speaks Out On Venues Lack Of Preparation For Washington Concert StrayKids 스트레이키즈 BangChan 방찬 https:t.co73MAIMuchV', 'metadata': {'total_chunks': 1, 'chunk_type': 'twitter', 'chunk_strategy': 'dynamic', 'chunk_index': 0, 'date': '2025-06-25T09:09:04', 'title': 'HypeMY', 'original_id': '1937800197234676085'}, 'distance': 0.7706171274185181, 'segment': 'community'}, {'source': 'twitter_embeddings', 'document': 'nowtoronto Sundays Stray Kids concert marked the debut show at the temporary outdoor venue, Rogers Stadium, where an estimated 50,000 people attended. Did you see the show? RogersStadium StrayKids Toronto https:t.co9cGvcNMY9w', 'metadata': {'chunk_strategy': 'dynamic', 'title': 'nowtoronto', 'original_id': '1939732925685125227', 'date': '2025-06-30T17:09:02', 'chunk_type': 'twitter', 'total_chunks': 2, 'chunk_index': 0}, 'distance': 0.7801396250724792, 'segment': 'community'}, {'source':
```


Time Boxing

 Bucket documents into time bins

 Compute avg sentiment per bin

 Detect threshold spikes

 Summary created

Threshold

Spike -> Summarize

Narrative

- Narrative takes trends and spikes from summary agent
- Creates a recommendation and actionable options

Notable Person Analyzer

LLM Cache Mode

☒ Use Cache

☐ Clear Cache

Enter some text

How was the Washington DC Straykids concert on June 23?

Demo



Notable Person Analyzer 🎤 🦸 🌀

LLM Cache Mode

☒ Use Cache

☐ Clear Cache

Enter some text

How was the Washington DC Straykids concert on June 23?

Summary Narrative/Recommendation Docs

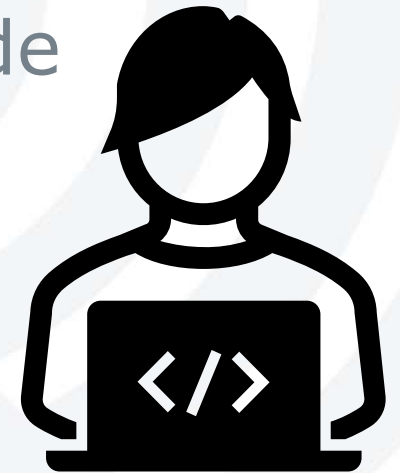
Takeaway: The June 23rd concert by Stray Kids at Nationals Park in Washington D.C., raised significant safety concerns due to extreme heat and staff unpreparedness, while some fans praised the group's response.

Spike: On stage during their June 23 performance, Stray Kids experienced a sentiment spike after mid-concert pauses were made mandatory following multiple heat-related emergencies requiring medical assistance. The concert also saw an unexpected negative sentiment spike when Bang Chan directly addressed the venue lack of air conditioning and staff preparation issues in his monologue.

Community: Fans reacted strongly to reports about the June 23rd Stray Kids concert's safety conditions; some defended the group while others criticized management regarding potential heat emergencies during their outdoor performance. Details unclear about specific fan experiences or how different members handled these concerns differently (Bang Chan seemed candid, David more positive)

Future Work

- Agentic data collection
 - Creating an agent that automatically collects required data
- Triggered expansion of corpus of data to include new entities
- Display of original OSINT artifacts



Why "Clarity" for Intelligence

- Utilizes Technological advancements to advance analytics
- Reveals connections previously missed
- Anticipatory recommendations for a fast-moving adversary
- Connecting disparate entities and events to reduce time to discover actionable content



Matthew Lessler
UMBC
B.S. Information Systems
Management
Class of 2027
mlessle1@umbc.edu



Julia Calvert
UMD – College Park
B.S. Information Science
Class of 2026
jcalver5@umd.edu



Quinn Dunnigan
Virginia Tech – Class of 2025
B.S. Cybersecurity and
Management Analytics
dquinn20@vt.edu



Jonathan Mei
UMD – College Park
B.S. Computer Science
Class of 2026
jmei2657@umd.edu



Izaiah Davis
UMGC – Class of 2025
B.S. Cybersecurity and
Technology
ldavis32@umgc.edu



Kyle Simon
MPS Data Science & Cybersecurity
M.Ed. Teaching & Curriculum
ksimon1@umbc.edu



Juliana Ochs

MICA – Baltimore
Class of 2026
B.A. UX/UI Design
jochs@mica.edu



Sachin Ashok

UMD – College Park
Class of 2026
B.S. Computer Science
sashok@umd.edu



Alyssa Miller

James Madison University
Class of 2027
B.S. Computer Science,
Minor Biotechnology (ISAT)
mille9ak@dukes.jmu.edu

Bentley Cech

University of Maryland – Baltimore
County
Class of 2026
B.A. Philosophy,
Minor Computer Science
bcech1@umbc.edu



Myra Cropper

Carnegie Mellon University
M.S. ECE
Class of 2026
B.S. Computer Science
mcropper@andrew.cmu.edu



EVERWATCH

A Booz Allen. COMPANY