

Background Summarization of Event Timelines

Adithya Pratapa, Kevin Small, Markus Dreyer

CMU

Amazon



Carnegie Mellon University
Language
Technologies
Institute



Consider the following event about Michael Jackson

● June 2009

Dr. Conrad Murray finds Jackson unconscious in the bedroom of his Los Angeles mansion. Paramedics are called to the house while he performs CPR, according to a recording of the 911 call. Murray travels with the singer in an ambulance to UCLA medical center where Jackson later dies.



June 2009

Feb 2010

Dr. Murray is charged with involuntary manslaughter. He pleads not guilty. Released on \$75,000 bail, he is allowed to resume practicing medicine, but the judge bans him from administering anesthetic agents, “specifically propofol”.



Michael Jackson's father, Joseph, files a wrongful death lawsuit against Murray.



Dr. Conrad Murray is sentenced to four years in county jail. Judge Michael Pastor says the evidence showed him to be guilty of a “continuous pattern of lies and deceit”.

Context to Events

- Popular events like MJ's death (2009—11), BP oil spill (2010—12) are long-running, covered extensively by global media outlets
- As a news reader, it can be challenging to keep track of these events
- Readers require necessary context to understand updates in these events

Consider the Feb 2010 update

● June 2009

● Feb 2010

Dr. Murray is charged with involuntary manslaughter. He pleads not guilty. Released on \$75,000 bail, he is allowed to resume practicing medicine, but the judge bans him from administering anesthetic agents, “specifically propofol”.

Who is Dr. Murray?

How is he related to the MJ's death?

● June 2009

● Feb 2010

Dr. Murray is charged with involuntary manslaughter. He pleads not guilty. Released on \$75,000 bail, he is allowed to resume practicing medicine, but the judge bans him from administering anesthetic agents, “specifically propofol”.

What is propofol?

Why was he on trial?

Background Summaries

- Background summary provides sufficient historical context to news updates
- We study backgrounds in the context of event timelines
- Related works: TV series recaps (Chen et al., 2022), scientific paper context (Hayashi et al., 2020), background knowledge (Louis, 2014; Peyrard and West, 2020)

New Dataset for Background Summarization

Train	
Swine Flu	2009
Financial crisis	2008
Iraq war	2005

Validation	
Haitian earthquake	2010
Michael Jackson death	2009—11
BP oil spill	2010—12

Test	
NSA leak	2014
Gaza conflict	2014
MH370 flight disappearance	2014
Yemen crisis	2011—12
Russian-Ukraine conflict	2014
Libyan crisis	2011
Egyptian crisis	2011—13
Syrian crisis	2011—13

Sources for timelines,

- [Timeline17 \(Binh Tran et al., 2013\)](#)
- [Crisis \(Tran et al., 2015\)](#)
- [SocialTimeline \(Wang et al., 2015\)](#)

Let's revisit the Michael Jackson event.

● June 2009

● Feb 2010

Dr. Murray is charged with involuntary manslaughter. He pleads not guilty. Released on \$75,000 bail, he is allowed to resume practicing medicine, but the judge bans him from administering anesthetic agents, “specifically propofol”.

June 2009

Feb 2010

Update

Dr. Murray is charged with involuntary manslaughter. He pleads not guilty. Released on \$75,000 bail, he is allowed to resume practicing medicine, but the judge bans him from administering anesthetic agents, “specifically propofol”.

Background

On June 25th, 2009, Dr. Conrad Murray found pop star Michael Jackson unconscious in the bedroom of his Los Angeles mansion. ... later died at UCLA medical center. ... A cocktail of drugs was found in the singer’s body, whose death by “acute propofol intoxication” has been ruled a homicide. Court documents revealed that Dr. Murray bought five bottles of propofol in May 2009 around the same time he was hired by Jackson. ...

Benchmark

Task setup

Systems

Metrics

Source: past updates

Target: background

(optional)

Query: current update

Flan-T5-XL

GPT-3.5 (zero-shot)

ROUGE

QuestEval

BERTScore precision

Results

	ROUGE-1	ROUGE-2	ROUGE-L	QuestEval	BERTScore Precision
Flan-T5-XL	41.4	17.4	37.6	25.0	85.6
GPT-3.5 (zero-shot)	37.7	11.7	33.0	30.5	87.2

Generic summarization

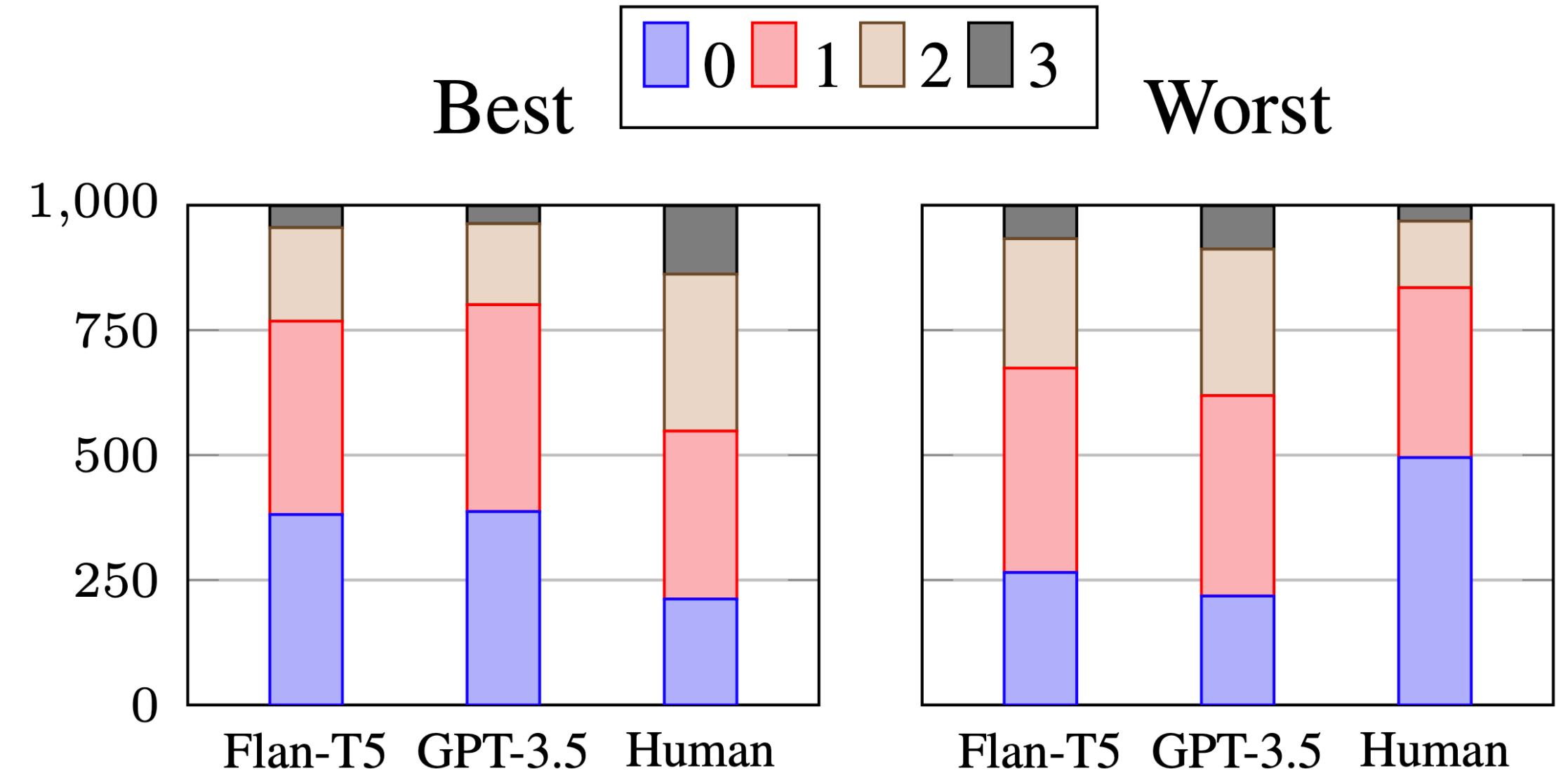
Results

	ROUGE-1	ROUGE-2	ROUGE-L	QuestEval	BERTScore Precision
Flan-T5-XL	41.3	17.4	37.6	24.9	85.6
GPT-3.5 (zero-shot)	40.5	12.9	35.9	31.7	87.5

Query-focused summarization

Human Evaluation

- Best-Worst Scaling (BWS) on MTurk
- Human >> Flan-T5 > GPT-3.5
- Turker's preferences show significant variance



Evaluating Backgrounds

- Standard metrics (ROUGE, BERTScore, QuestEval) don't evaluate the 'utility' of a background
- We propose a metric, Background Utility Score (BUS)

Background Utility Score (BUS)

● Apr 20 2010

● Apr 24 2010

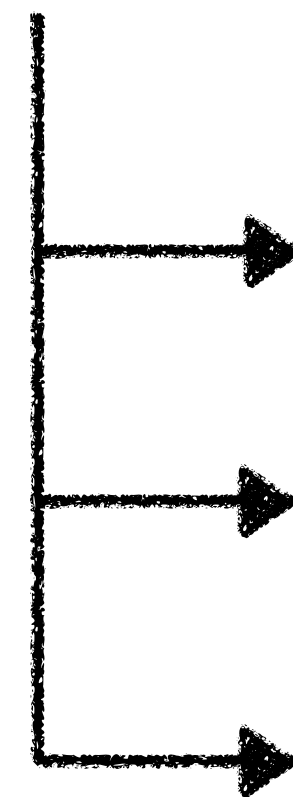
For the first time, oil is found to be leaking from the well. Pressure and release rate are unknown. A homeland security report on critical infrastructure says the problem has “no near-term impact to regional or national crude oil or natural gas supplies.”

Consider the event, BP oil spill

Background Utility Score (BUS)

● Apr 24 2010

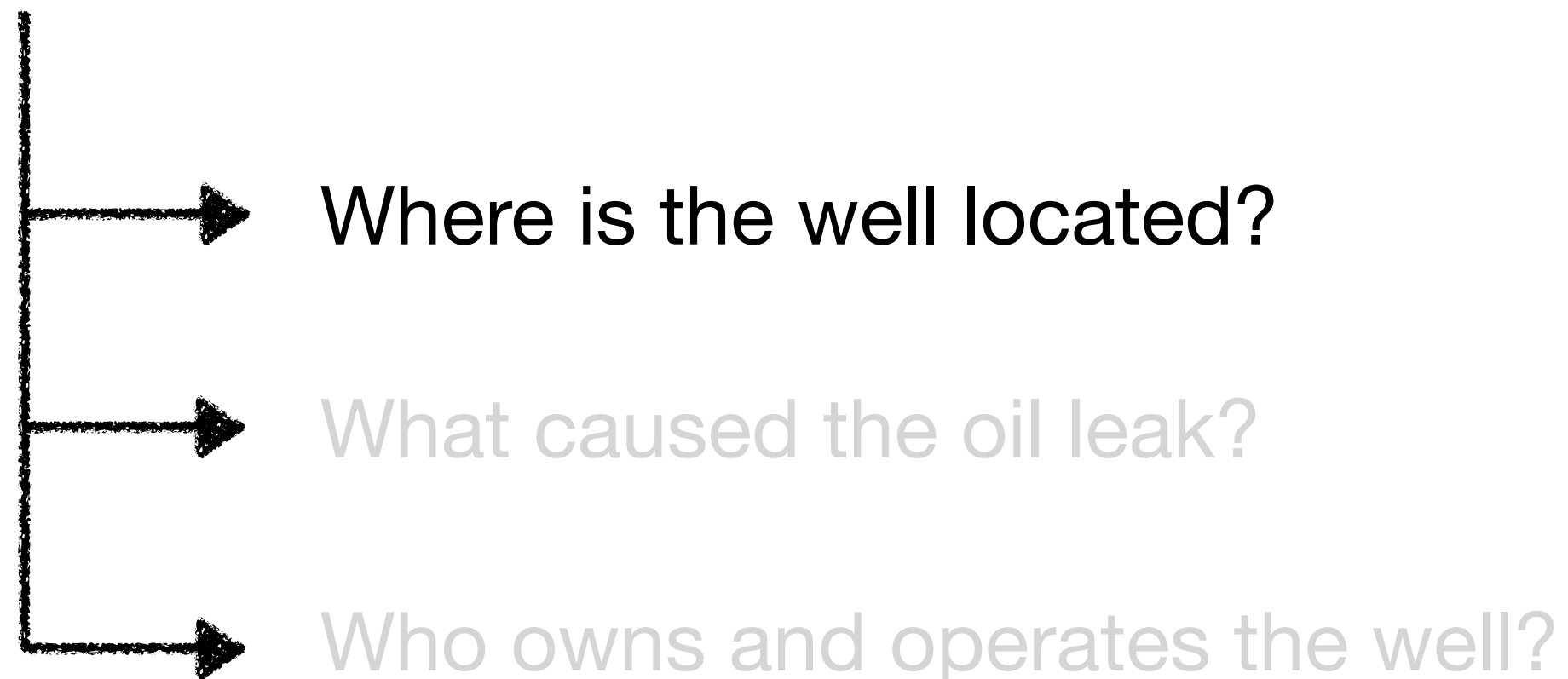
For the first time, oil is found to be leaking from the well. Pressure and release rate are unknown. A homeland security report on critical infrastructure says the problem has “no near-term impact to regional or national crude oil or natural gas supplies.”

- 
- Where is the well located?
 - What caused the oil leak?
 - Who owns and operates the well?

Background Utility Score (BUS)

Apr 24 2010

For the first time, oil is found to be leaking from the well. Pressure and release rate are unknown. A homeland security report on critical infrastructure says the problem has “no near-term impact to regional or national crude oil or natural gas supplies.”



[Flan-T5-XL] The well is located in the Gulf of Mexico.

[GPT-3.5] The well is located in the Gulf of Mexico.

[Human] The well is located 5,000 ft below the surface of the Gulf of Mexico, 84km south-east of Venice, Louisiana.

Background Utility Score (BUS)

● Apr 24 2010

For the first time, oil is found to be leaking from the well. Pressure and release rate are unknown. A homeland security report on critical infrastructure says the problem has “no near-term impact to regional or national crude oil or natural gas supplies.”

- Where is the well located?
- What caused the oil leak?
- Who owns and operates the well?

[Flan-T5-XL] The oil started leaking due to the explosion of the Deepwater Horizon drilling rig.

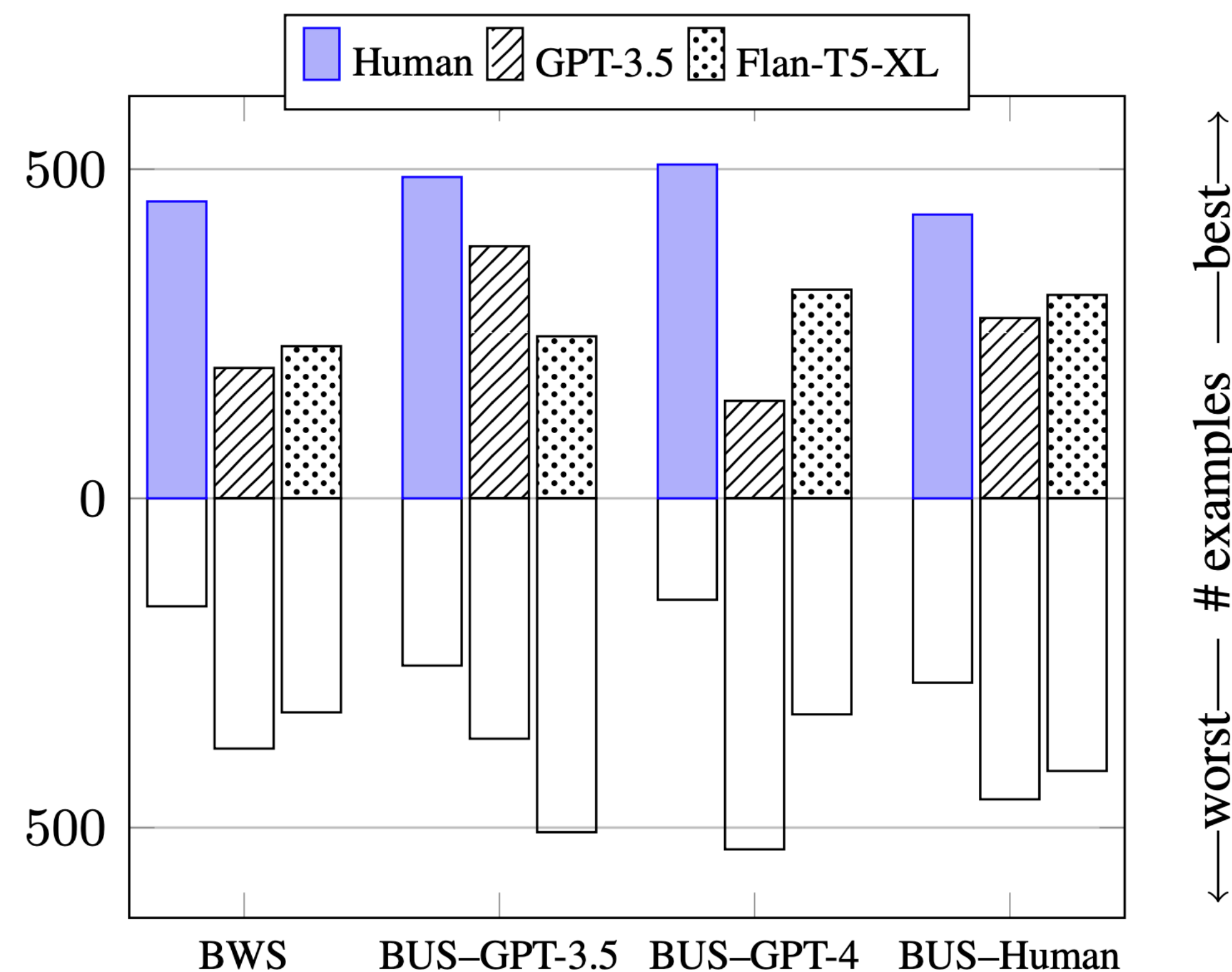
[GPT-3.5] N/A

[Human] N/A

BUS results

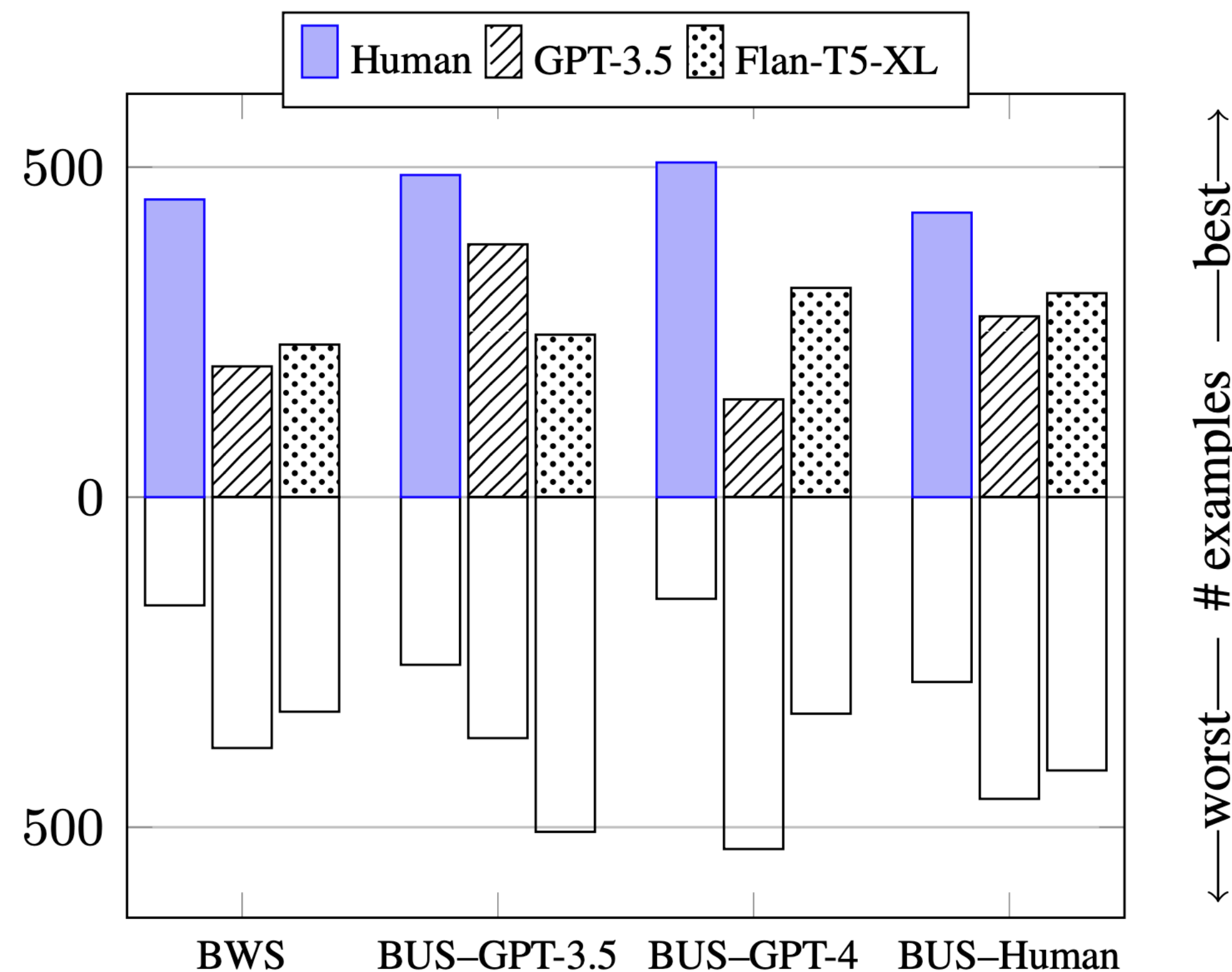
For 1k test samples, we rank Flan-T5, GPT-3.5 and human backgrounds on a best-worst scale

- BWS: MTurk for directly comparing backgrounds
- BUS-★: use BUS to rank backgrounds



BUS results

- BUS—GPT-4 aligns better than GPT-3.5 with our human evaluation (BWS)
- Also simulate BUS—Human, with Q&A pairs compiled from MTurk
- BUS—human aligns well standard human evaluation (BWS)



Summary

- A new task: background summarization of events
- A new expert-annotated dataset covering 14 events
- A benchmark of Flan-T5, LongT5, and GPT-3.5
- An QA-based metric to measure utility of a background
- Future directions: personalized backgrounds, local events, expanded sources

Paper



Code & Dataset

