Explicit, Implicit, and Scattered: Revisiting Event Extraction to Capture Complex Arguments

Omar Sharif, Joseph Gatto, Madhusudan Basak, Sarah M. Preum

Persist Lab, Department of Computer Science

DARTMOUTH

Motivating Example: Event Understanding on Social Media

Goal: Aggregate self-reported information from Cancer Patients on Reddit for Public Health Researchers

Information-seeking post

I'm 46 and was recently diagnosed with prostate cancer... My doctor wants to go the da vinci route, but I'm not sure I want to do that....If I am being honest, I really want to try other therapies such as FLA. ... Sadly, my cancer has spread and I am running out of options...Did any of you had multi-focal cancer and have had success longer term with radiation?

Social Media Post

Target Information

- How Old Is The Patient?
- What Type of Cancer Do They Have?
- What Types of Treatments Do They Recieve?
- What is the status of their cancer?

Motivating Example: Event Understanding on Social Media

Goal: Aggregate self-reported information from Cancer Patients on Reddit for Public Health Researchers

Information-seeking post

I'm 46 and was recently diagnosed with prostate cancer... My doctor wants to go the da vinci route, but I'm not sure I want to do that....If I am being honest, I really want to try other therapies such as FLA. ... Sadly, my cancer has spread and I am running out of options...Did any of you had multi-focal cancer and have had success longer term with radiation?

Social Media Post

Target Information

- **How Old Is The Patient?**
  46
- **What Type of Cancer Do They Have?**
  Prostate Cancer
- **What Types of Treatments Do They Recieve?**
  Prostate removal surgery
- **What is the status of their cancer?**
  Multi-focal and metastasized

Motivating Example: Event Understanding on Social Media

Goal: Aggregate self-reported information from Cancer Patients on Reddit for Public Health Researchers

Information-seeking post

I'm 46 and was recently diagnosed with prostate cancer... My doctor wants to go the da vinci route, but I'm not sure I want to do that....If I am being honest, I really want to try other therapies such as FLA. ... Sadly, my cancer has spread and I am running out of options...Did any of you had multi-focal cancer and have had success longer term with radiation?

Social Media Post

Structured Representation through Event Arguments

Event Type: Cancer	
Roles	**Arguments**
Age	46
Cancer type	Prostate cancer
Treatment	Prostate removal surgery
Cancer Status	Multi-focal and metastasized

Limitations & Motivation

❌ Existing works extracts arguments as ***a span in the text*** which is limiting,

- ***Implicit:*** not directly mentioned in the text but can be inferred through context.
- ***Scattered:*** composed of information scattered throughout the text.

Information-seeking post

I'm 46 and was recently diagnosed with prostate cancer... My doctor wants to go the da vinci route, but I'm not sure I want to do that....If I am being honest, I really want to try other therapies such as FLA. ... Sadly, my cancer has spread and I am running out of options...Did any of you had multi-focal cancer and have had success longer term with radiation?

Example demonstrating complex event arguments in online discourse

Explicit Arguments: ✅

- *Age*: 46
- *Cancer Type*: prostate

Implicit Arguments: ❌

- *Gender*: male
- *Treatment Option*: prostate removal surgery

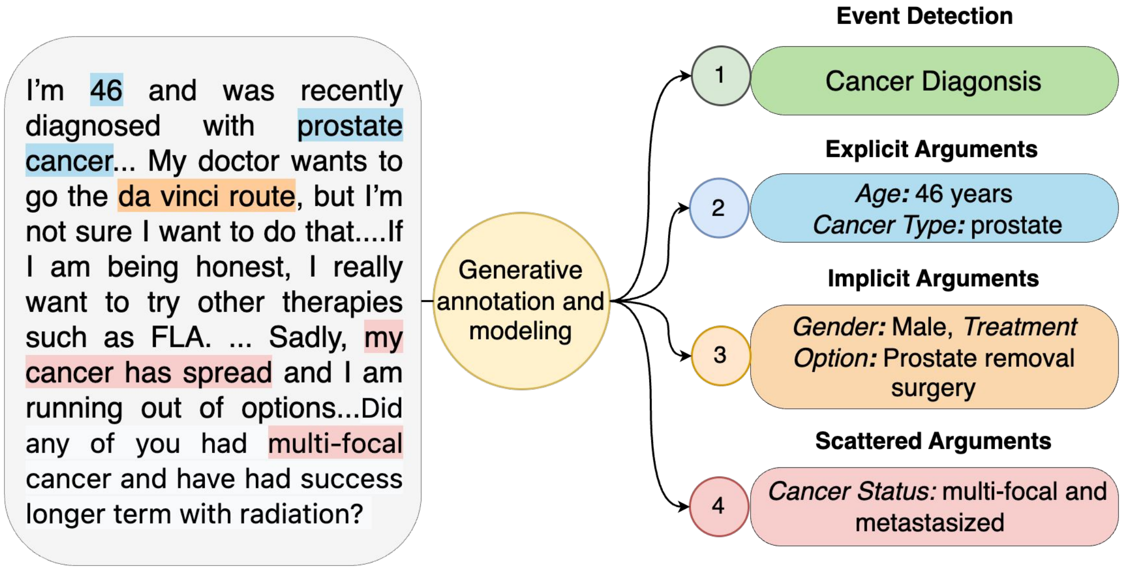
Scattered Arguments: ❌

- *Cancer Status*: Multifocal and metastasized

Contribution: Rethinking EE Annotation as Generation

🧠 We reformulate the ***annotation as generation*** where the arguments written as ‘free text’.

👍 This approach offer flexibility on 4 event extraction tasks (***detection, explicit, implicit, scattered***).



Event extraction annotation as text generation

Contribution: Novel Discourse-level Event-Extraction Dataset

Dataset (year)	Explicit Arguments	Implicit Arguments	Scattered Arguments	Long Documents/ Discourse
ACE '04	✅	❌	❌	❌
RAMS '20	✅	❌	❌	❌
WikiEvents '21	✅	❌	❌	✅
DocEE '22	✅	❌	❌	✅
ArgGen '22	✅	❌	❌	✅
DiscourseEE '24	✅	✅	✅	✅

1. Doddington et al. (2004) "The Automatic Content Extraction (ACE) Program Tasks, Data, and Evaluation "
2. Ebner et al. (2020) "Multi-Sentence Argument Linking"
3. Li et al. (2021) "Document-Level Event Argument Extraction by Conditional Generation"
4. Tong et al. (2022) "DocEE: A Large-Scale and Fine-grained Benchmark for Document-level Event Extraction"
5. Zhang et a. (2024) "ArgGen: Prompting Text Generation Models for Document-Level Event-Argument Aggregation"

Contribution Summary

📝 Framework

➔ Utilized event-extraction framework for characterizing ***discourse-level health advice*** from ***social media*** data.

📚 Resource

➔ Developed a ***novel health-advice tailored event-ontology*** and *a* ***novel dataset (DiscourseEE)***.

🧑‍💻 Modeling

➔ Explore ***supervised training***, ***instruction fine-tuning*** and ***in-context learning*** settings.

DiscourseEE:
A Novel EE Resource

Annotating Complex Arguments

- To capture complexity of health-discourse we define 4 types of arguments,

Information-seeking post

I am a 42-year-old male with severe back pain. I haven't taken my 12 mg of suboxone since Thursday. My last opioid was 3 days ago. My nose has been runny for the last 2 days, and I feel like an 6/10. Will it kick in?

Advice-containing comment

Yeah it will kick in. Withdrawals are coming. Suboxone just has an extremely long half life which is why you are still feeling fine. It will catch up to you though. I definitely wouldn't recommend jumping off at 12mg!

Extracted arguments from the post-comment pair

Core Arguments

Subject/Patient: individual doesn't feel withdrawals after not taking suboxone
Tapering event: have not taken 12mg of suboxone since Thursday
Effects: runny nose and feeling uncomfortable

Type-specific Arguments

Argument roles	Values
Taper Condition	was on 12 mg suboxone
Trigger	stop taking suboxone
Tapering Type	self-tapering
Taper Medications	suboxone
Initial dosage	12mg
Current dosage	0mg
Goal dosage	0mg
Tapering start time	3 days ago
Target duration	null

Subject-specific Arguments

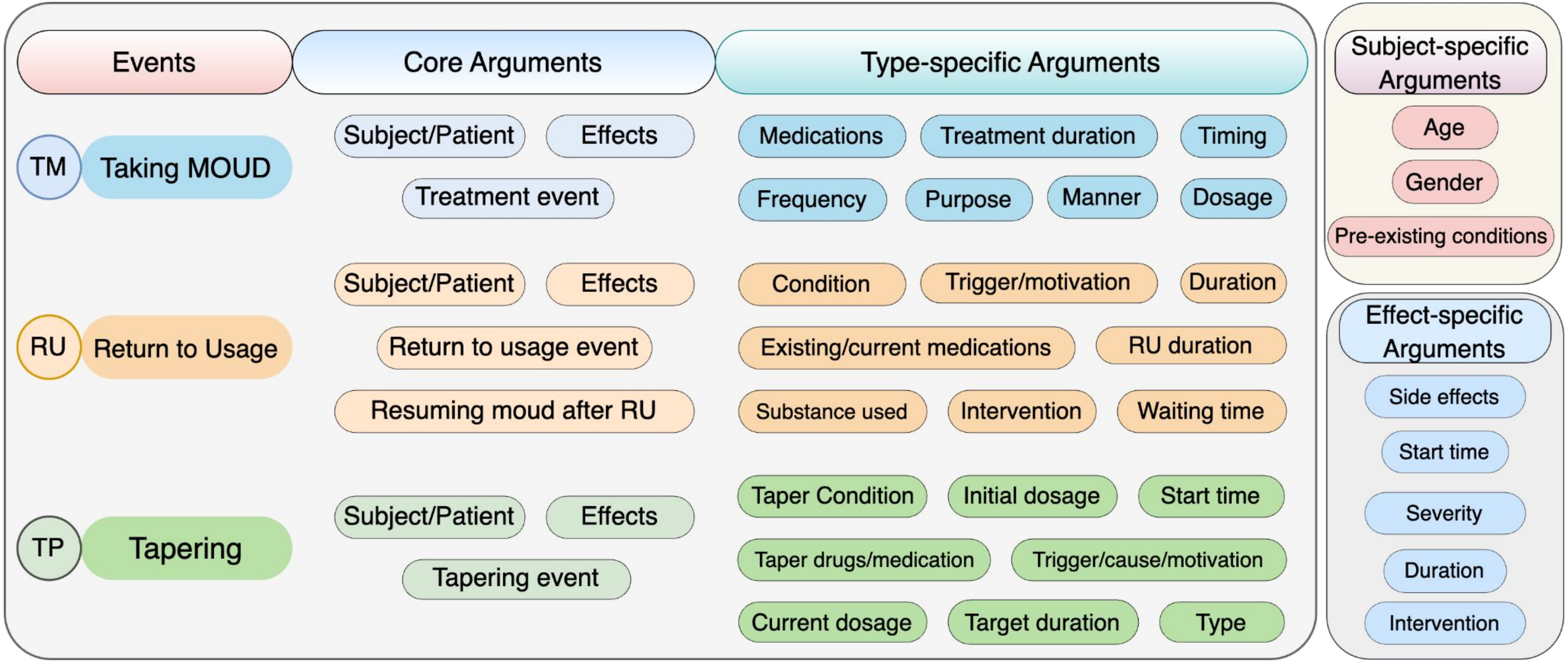
Argument roles	Values
Age	42 year
Gender	male
Pre-existing conditions	severe back pain

Effect-specific Arguments

Side effects	runny nose, withdrawals
Severity	medium
Side-effect start time	null
Side-effect duration	2 days
Intervention	not jumping off at 12mg

DiscourseEE Ontology

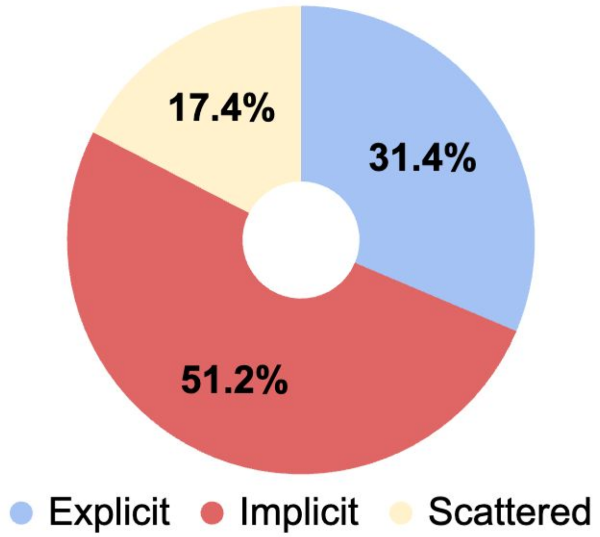
- We defined total ***41 arguments*** across three event types.



Event ontology of DiscourseEE dataset.

DiscourseEE Statistics

	Dataset
#Samples	396
Avg. sample length (#words)	**114.80**
#Arguments	3845
Avg. #arguments per doc	**10.02**

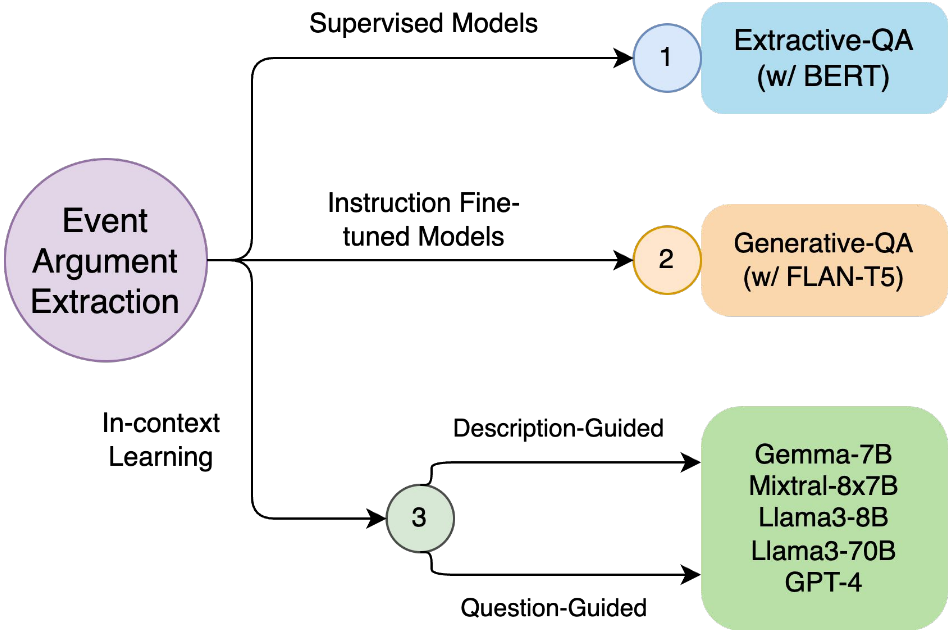


- Higher presence of **implicit** and **scattered** arguments in social discourse.
- DiscourseEE samples are longer with high argument density.

Benchmarking: Methods & Results

Event Argument Extraction Models

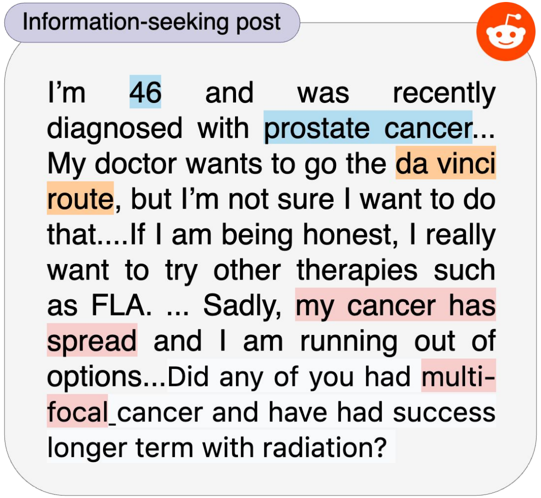
- Experimented event argument extraction (EAE) experiments in 3 distinct settings: ***Extractive-QA***, ***Generative-QA***, and ***LLM with varying prompts***



Event argument extraction experimental settings

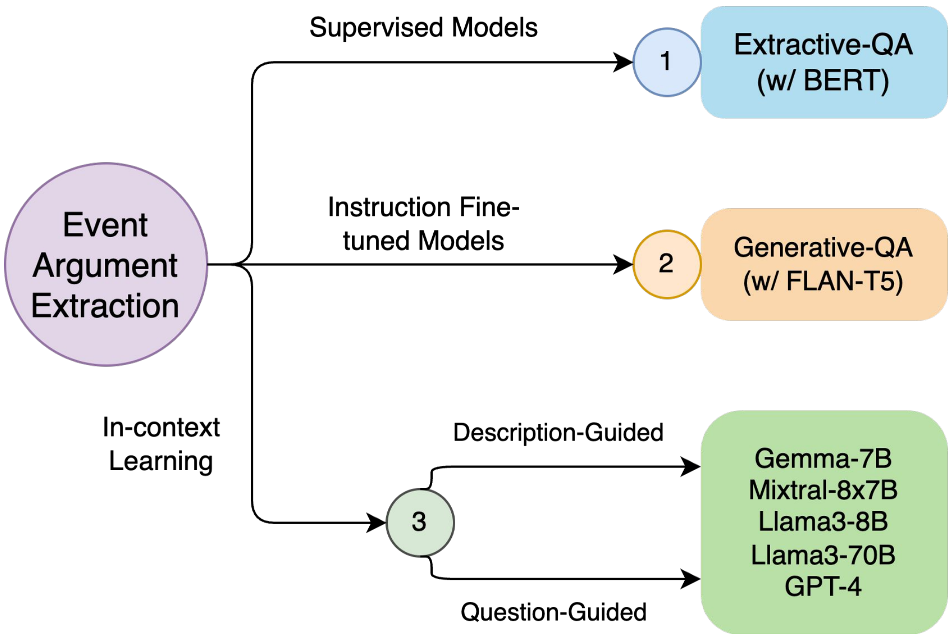
⭐ Description-Guided ⭐

Prompt: "Describe all treatments used in this post: "



Event Argument Extraction Models

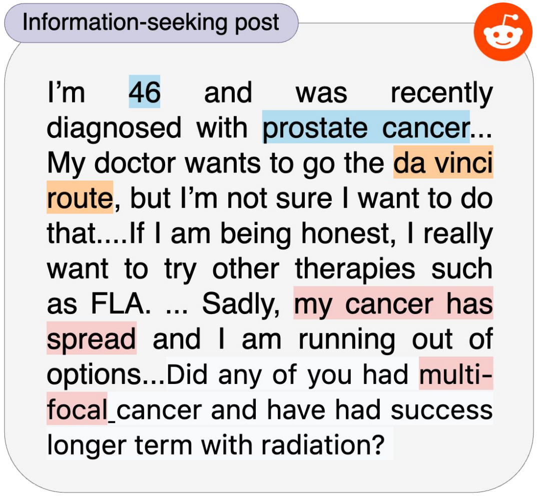
- Experimented event argument extraction (EAE) experiments in 3 distinct settings: ***Extractive-QA***, ***Generative-QA***, and ***LLM with varying prompts***



Event argument extraction experimental settings

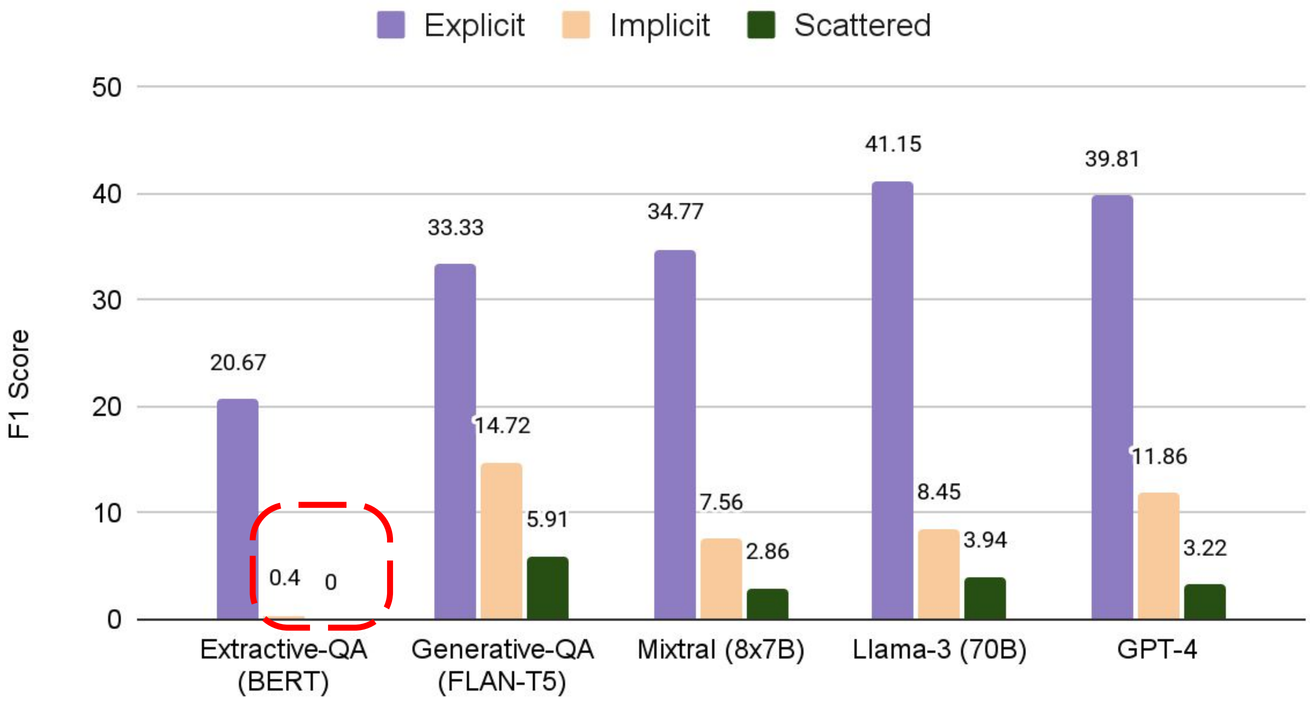
⭐ **Question-Guided** ⭐

Prompt: "What are the treatments used in this post?"



Argument Extraction in Exact Match Setting

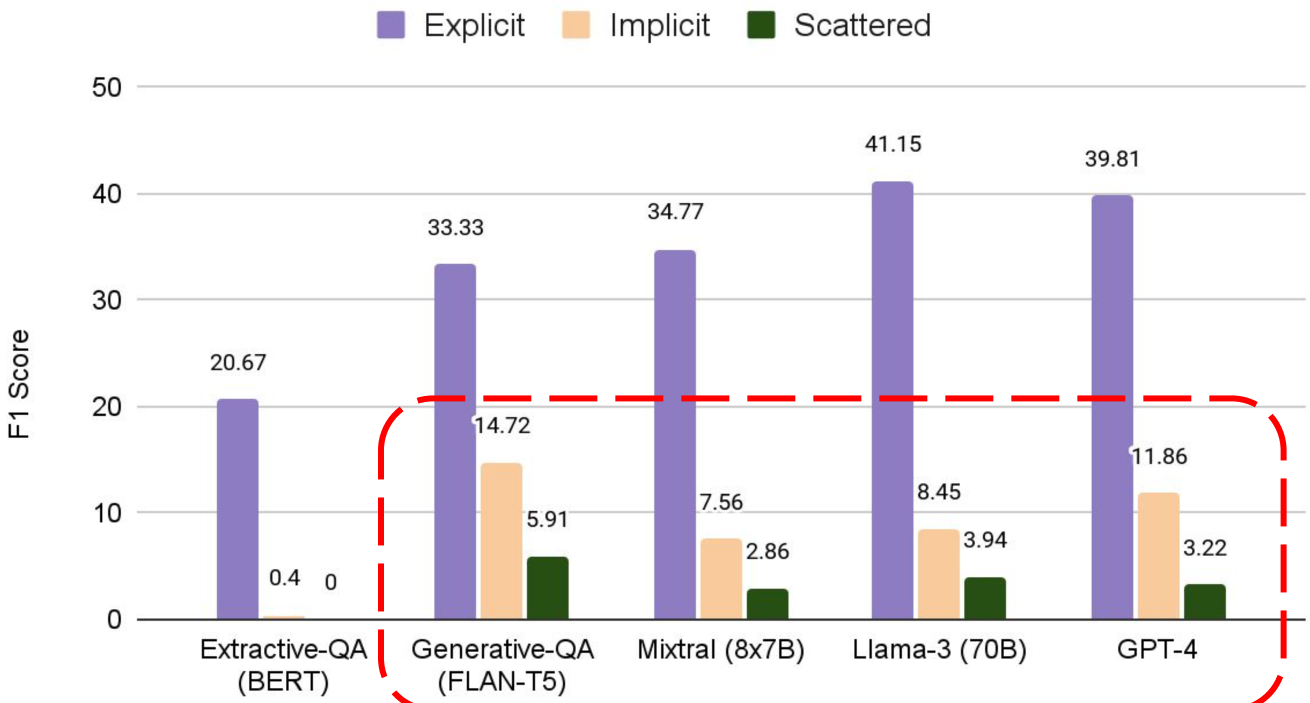
❌ Traditional extractive models can not capture implicit and scattered arguments.



Performance of baselines and best-performing LLMs under Exact Match

Argument Extraction in Exact Match Setting

💡 Exact match severely underestimates the performance of generative models.

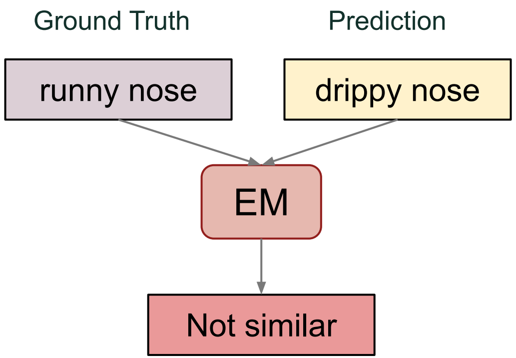


Performance of baselines and best-performing LLMs under Exact Match

Revisiting Evaluation

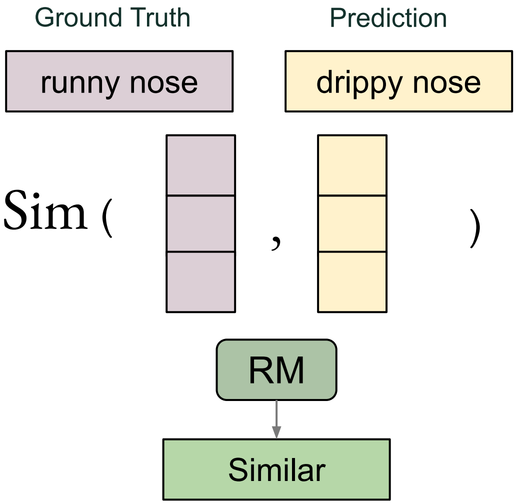
❌ Exact Match (EM)

- ***'Exact match'*** strictly match the start and end index for evaluation.
- Incorrectly consider valid outputs as wrong.



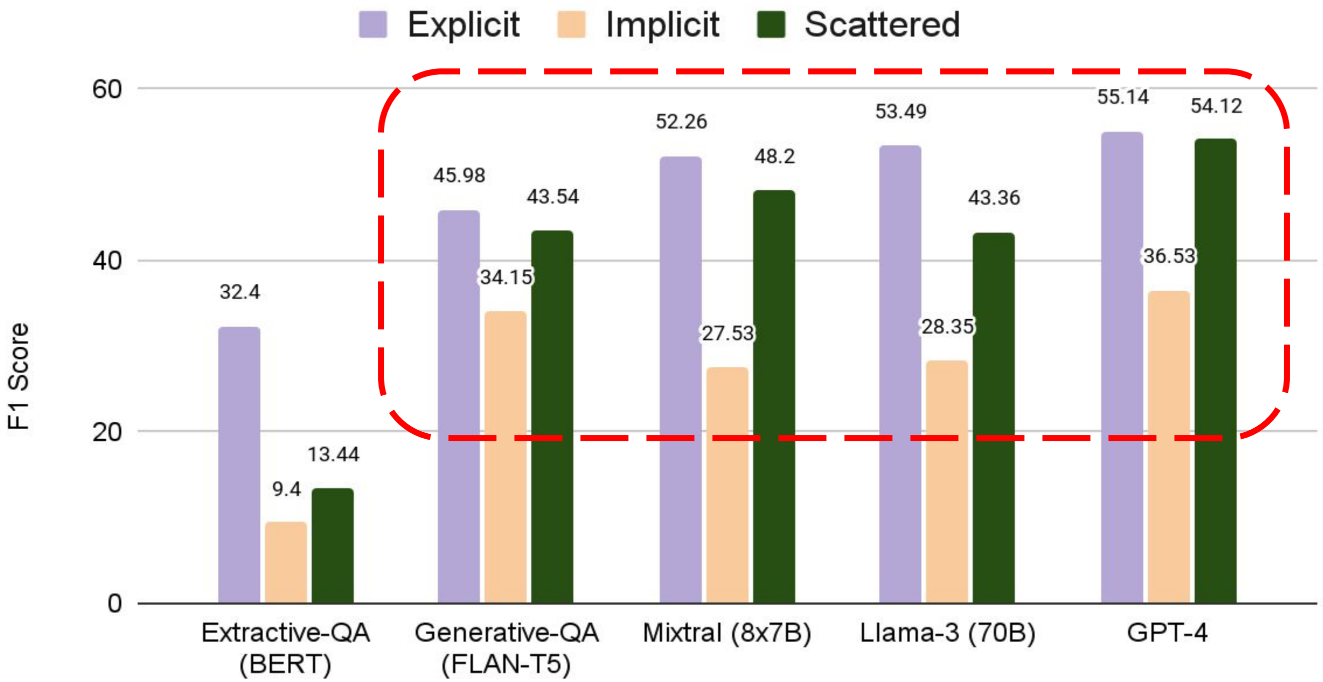
👍 Relaxed Match (RM)

- Can capture the variation of prediction from the ground truth.
- Relaxed match approach based on semantic similarity gives more accurate evaluation.



Argument Extraction in Relaxed Match Setting

- ✅ Performance improves across all models in relaxed-match
- 👎 Best performing generative models (i.e GPT-4, Llama-3) still struggles.



Performance of baselines and best-performing LLMs under Relaxed Match

Overall Argument Extraction Results

Performance comparison of baseline models and best-performing LLMs for argument extraction.

Observations

- Comparable performance by Instruction fine-tuned models.
- Llama3-70B obtained maximal score among open-source LLMs.
- GPT-4 achieved best performance but still struggled.

Future Directions

🧠 Improve ability of generative models to identify implicit and scattered arguments.

🔬 Further investigation of in-context learning techniques.

🧑‍💻 Improve evaluation approach.

Thank You!

Explicit, Implicit, and Scattered:
Revisiting Event Extraction to Capture Complex Arguments

Omar Sharif, Joseph Gatto, Madhusudan Basak, Sarah M. Preum

Code and data available at:
https://omar-sharif03.github.io/DiscourseEE/

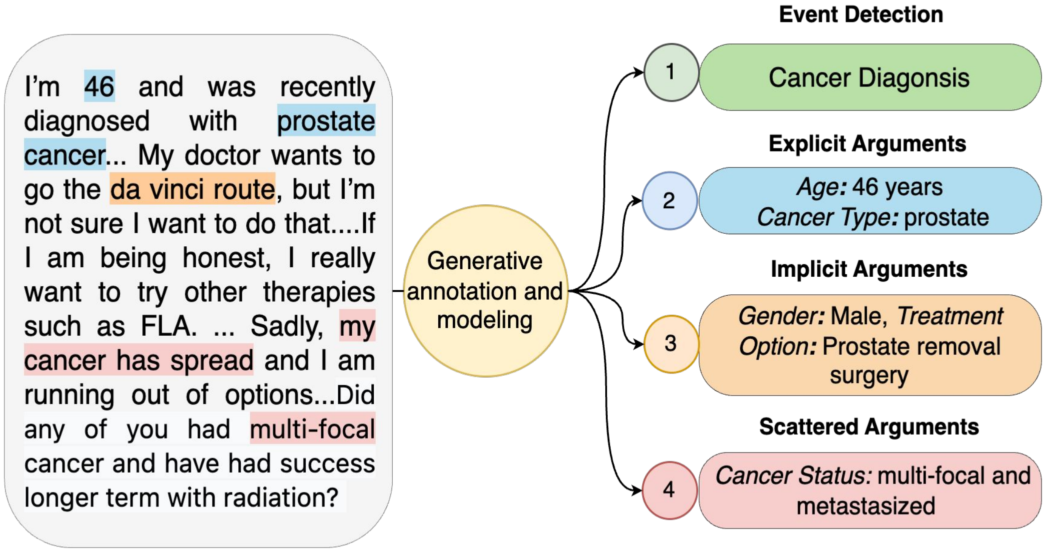
Recap

Key Idea: Rethinking event-extraction as text generation introducing and enabling extraction of *Implicit* and *Scattered arguments* from complex *Social Discourse*.

✅ Reformulating event extraction

✅ Novel Ontology and dataset

✅ Relaxed matching based evaluation and extensive benchmarking



Event Extraction as text generation