

From Information to Insight: Leveraging LLMs for Open Aspect-Based Educational Summarization



Yang Zhong¹ and Diane Litman^{1,2}

¹ Department of Computer Science, ² Learning Research and Development Center
University of Pittsburgh, Pittsburgh, PA, 15213



INPUT: All Reflections: (26 students)

- I found it interesting using the loop rule to solve the last problem because it made the problem more simple than I initially thought.
- I liked that no matter which path you took to find current in a circuit, you'll get the same value.
- I thought that the applications of grounding and how that could simplify complicated circuits was interesting
- Grounding and how it changes the calculation for current
- I found the part where the professor left the class as funny. It really gave us all a break in the middle
- Learning the paths circuits take and what way the current flows I found to be very interesting. It's cool to see how current flows in real wires and in real life.
- I found the videos explaining the concepts most interesting.
- The grounding was very interesting.
- I thought grounding was interesting. I never understood what grounding meant before.
- Current here split up when you add grounded points to a circuit
- Batteries and circuits are intriguing. I like to learn about currents and how energy within batteries work
- The circuit analysis about the voltage in loops summing to 0.
- I thought your explanation of ammeters and voltmeters was very well done ... how current runs through a wire and why each of these have to have very different resistances.
- How you can use instruments to measure voltage and current and how these instruments can affect these values
- I thought it was interesting that all grounded points are connected to each other.
- Voltmeters and ammeters are really interesting in that I think it's cool that we can actually measure these concepts (voltage and current) pretty easily.
- ...Omitted the remaining

Prompt: Please write a short summary with no more than 100 words, focusing on the topic of **Current** based on below reflections

Aspect: Current **Human-annotated Cluster** marked w/ ⚡

Summary:

Many students found the topic of current to be interesting. Students specifically mention *how current flows in real life* and how grounding changes calculation for current to be interesting. Students also found *using instruments such as ammeters and voltmeters to measure current* to be interesting, and *how these instruments can affect current values*.

Prompt: Please write a short summary with no more than 100 words, focusing on the topic of **Grounding** based on below reflections

Aspect: Grounding **Human-annotated Cluster** marked w/ ★

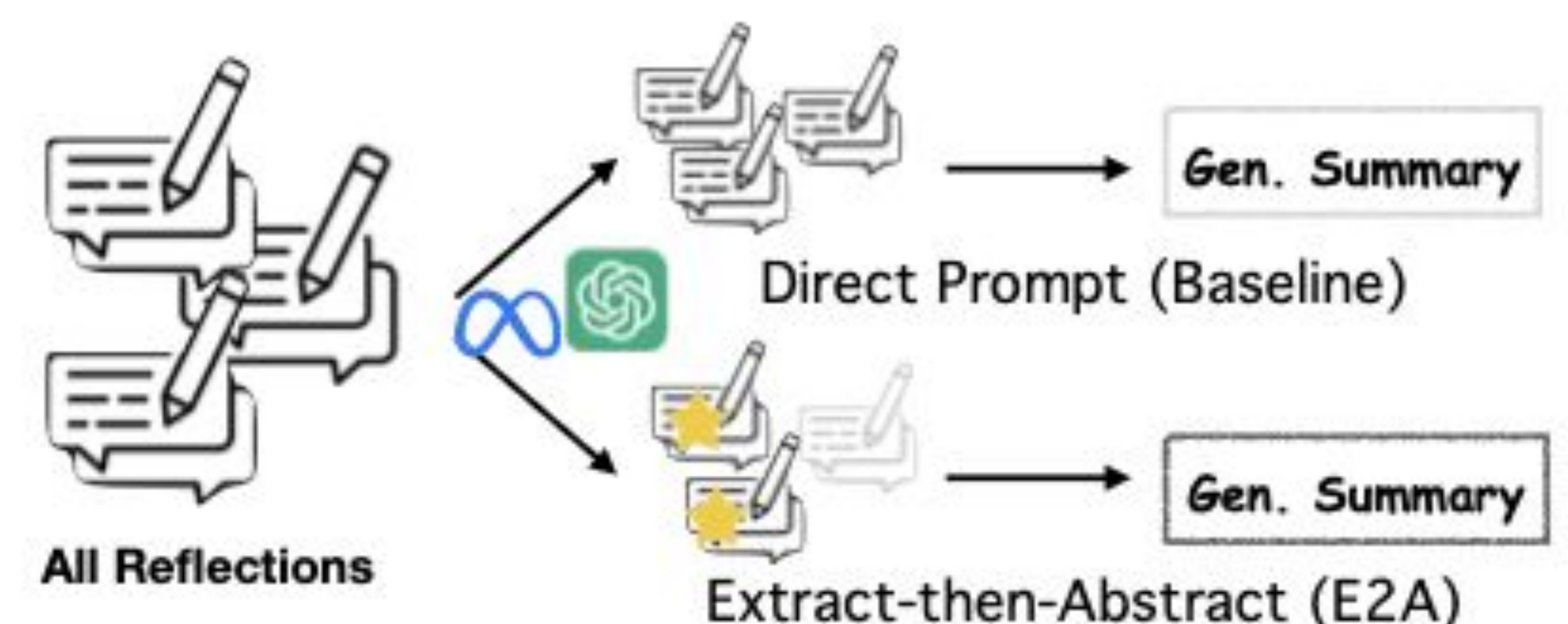
Summary:

Grounding was a topic that many students found interesting in today's lecture. Students also *enjoyed how grounding simplified calculations*. Students enjoyed seeing the concept apply to *circuit analysis*. Finally, *the idea that all grounded points are connected to each other* was found to be interesting by several students.

Prior Work (ReflectSUMM)
(general abstractive summary)

Prompt: Given student's responses, create a short summary using your own words.

The students today found circuits and its components really interesting. They also found ammeters and voltmeters extremely interesting with their applications. The students also found the concept of grounding and current really interesting, and a couple of students found the loop rule interesting.



Experimental setup and Evaluation Options

Examples from our ReflectASP dataset and a comparison to prior work

Motivation of Open Aspect-based Summarization of Reflections

- Diverse aspects can vary across disciplines and text dependent.
- There is a **lack of studies in the educational domain**, which helps avoid data contamination due to the exclusive curation from real-world scenarios.

Research Questions

RQ1. How well do large language models (LLMs) generate aspect-based summaries of reflections in a zero-shot setting?

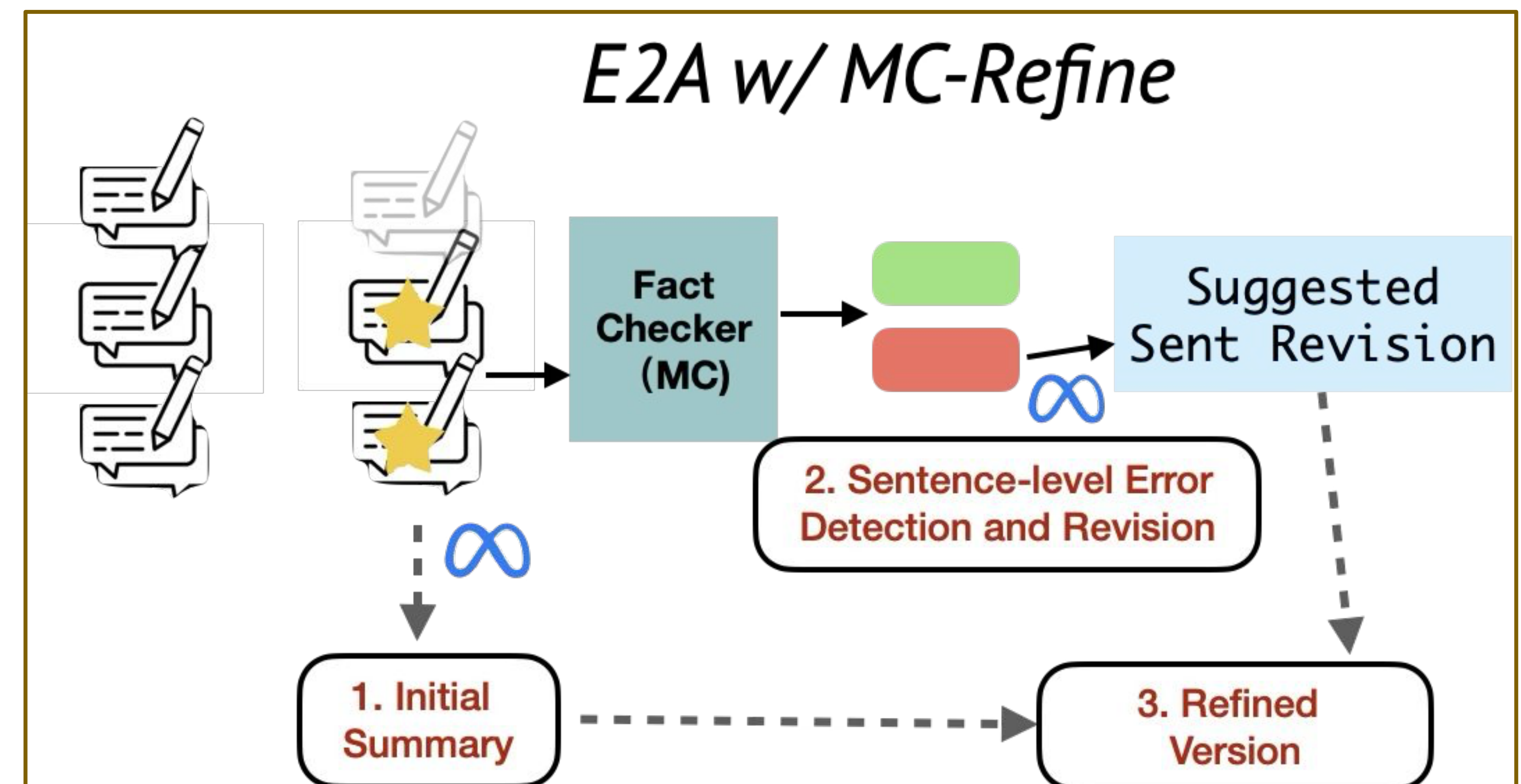
RQ2. How does different refinement help with the summarization?

Dataset

- Student reflections on lectures collected from real-world lectures
- 313** input-summary pairs
 - Unique **manually** annotated clusters of aspect-related reflections
 - Human-edited summaries** on top of GPT-4 generated versions

Experimental Setup

- Two** types of generation: Direct Prompting Extract-then-Abstract (E2A)
- Four** Sets of Evaluation Metrics
- Multiple** LLM backbones and **Three** Refinement Strategies

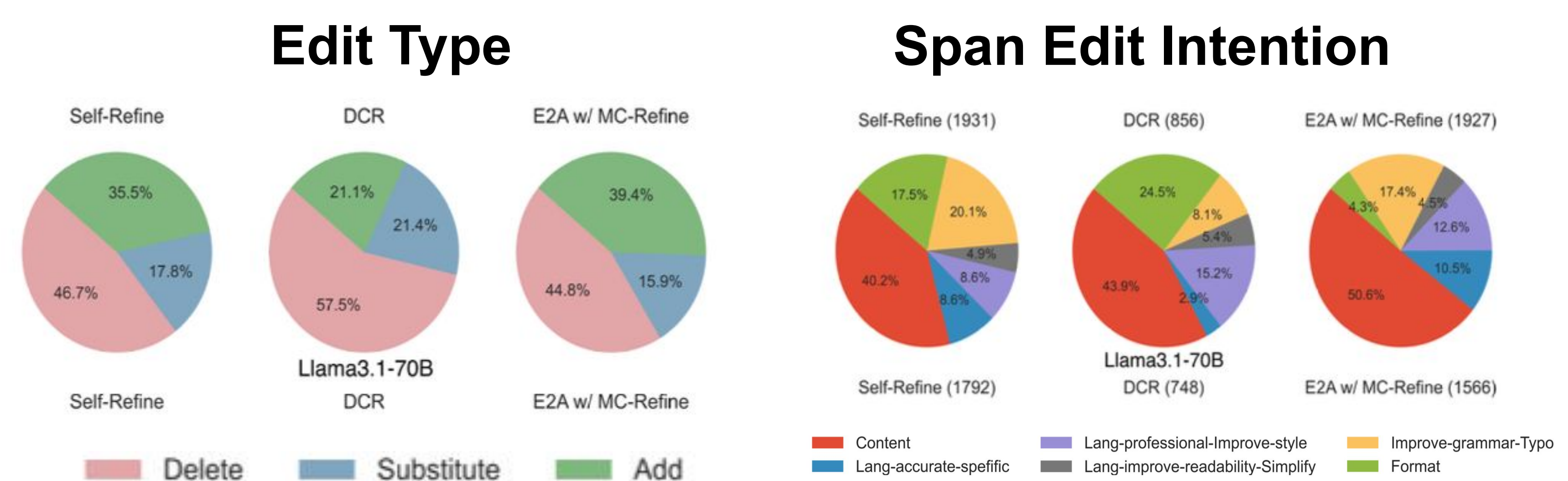


Result

- E2A improved on top of baseline direct prompting
- Our refinement strategy improves factuality and overall quality, as confirmed by both automatic and human evaluations

ID	Model	R-1	R-2	R-L	BS	MC _{EXT}	MC _{INPUT}	# Sents	# Words
	LLAMA3.1-70B	47.54	18.64	42.12	90.04	53.97	89.48	3.55	94.9
	w/ Self-Refine	46.05	16.88	40.37	89.90	55.74*	86.58	2.93	82.6
	w/ DCR	38.25	14.08	34.22	89.89	71.08	91.77	2.45	42.2
	E2A	48.75*	18.79*	43.17*	90.08	69.55*	88.20	3.51	84.7
	w/ DCR	34.84	13.47	31.25	89.83	82.43	93.37	1.94	32.1
	w/ MC-Refine	47.80	18.44	42.40	90.34*	71.79*	90.50*	3.43	79.9
Proprietary LLMs									
	GPT3.5-turbo	35.72	9.95	32.05	88.04	44.85	91.25	4.96	100.8
	GPT4	33.51	7.08	29.06	87.50	51.30	89.82	4.05	97.3
	GPT4o	34.15	7.93	30.48	87.52	58.07	91.57	4.89	102.9
	Human (Oracle)		N/A			76.97	87.42	3.49	69.2

Revision Analysis



Our proposed approach (E2A w/ MC-Refine) exhibits diverse edits, while DCR tends to delete excessively

Our proposed approach (E2A w/ MC-Refine) performs more edits, emphasizing content updates and enhanced specificity