Explanatory Summarization with Discourse-Driven Planning

Dongqi Liu[Ω], Xi Yu[Ω], Vera Demberg[Ω], Mirella Lapata[ϴ]

[Ω]Saarland University, Germany

[ϴ]University of Edinburgh, The United Kingdom

Accepted by the Trans. of the Association for Computational Linguistics (TACL 2025)

ACL 2025 VIENNA

Introduction

Lay Summarization Challenge:

- Scientific texts need **explanatory content** for accessibility
- Current methods **underrepresent** explanations in generated summaries

Key Insights:

- Explanations help lay readers understand complex concepts through **analogies, background, comparisons**
- Most models follow end-to-end approaches **without explicit explanation modeling**
- Plan-based models offer better **controllability** and reduce **hallucinations**

Method

Step 1: Explanatory Content Extraction

- Apply `DMRST` parser to identify explanatory EDUs from RST relations
- Focus on 4 key relations: *Background*, *Elaboration*, *Explanation*, *Comparison*
- Extract (explanation, target) EDU pairs from reference summaries

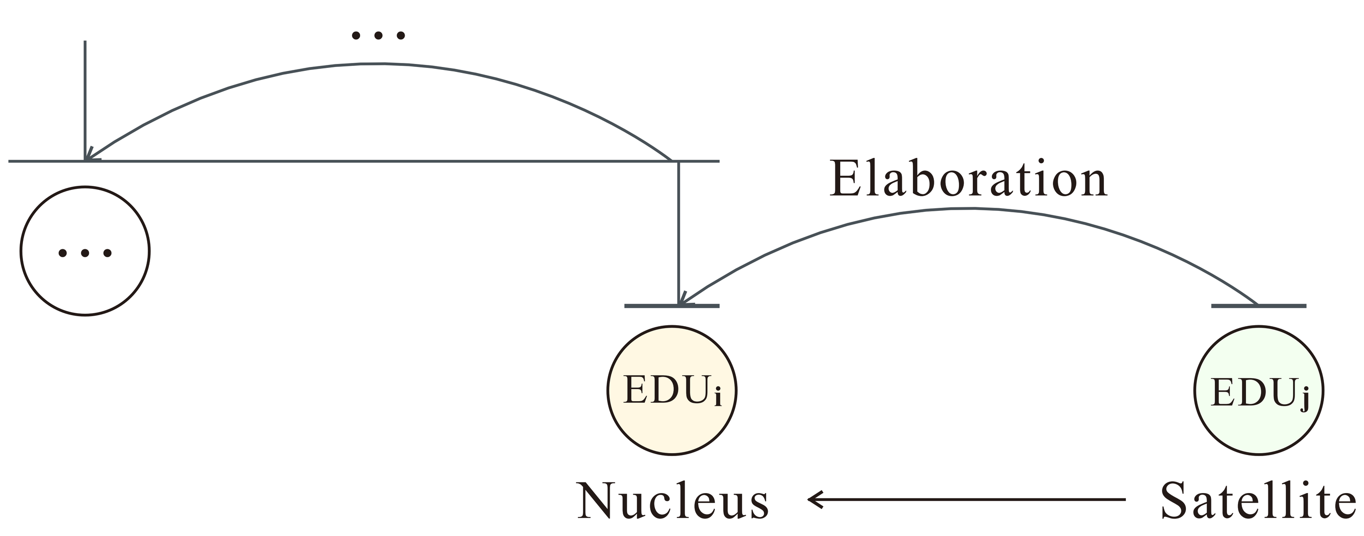

Figure 1. RST tree structure

Step 2: Plan Generation

- Use `GPT-4o` to generate questions from target sentences + context
- Questions trigger explanatory content (based on QUD framework)
- Create ordered sequence of plan questions $b = (q_1, q_2, ..., q_n)$

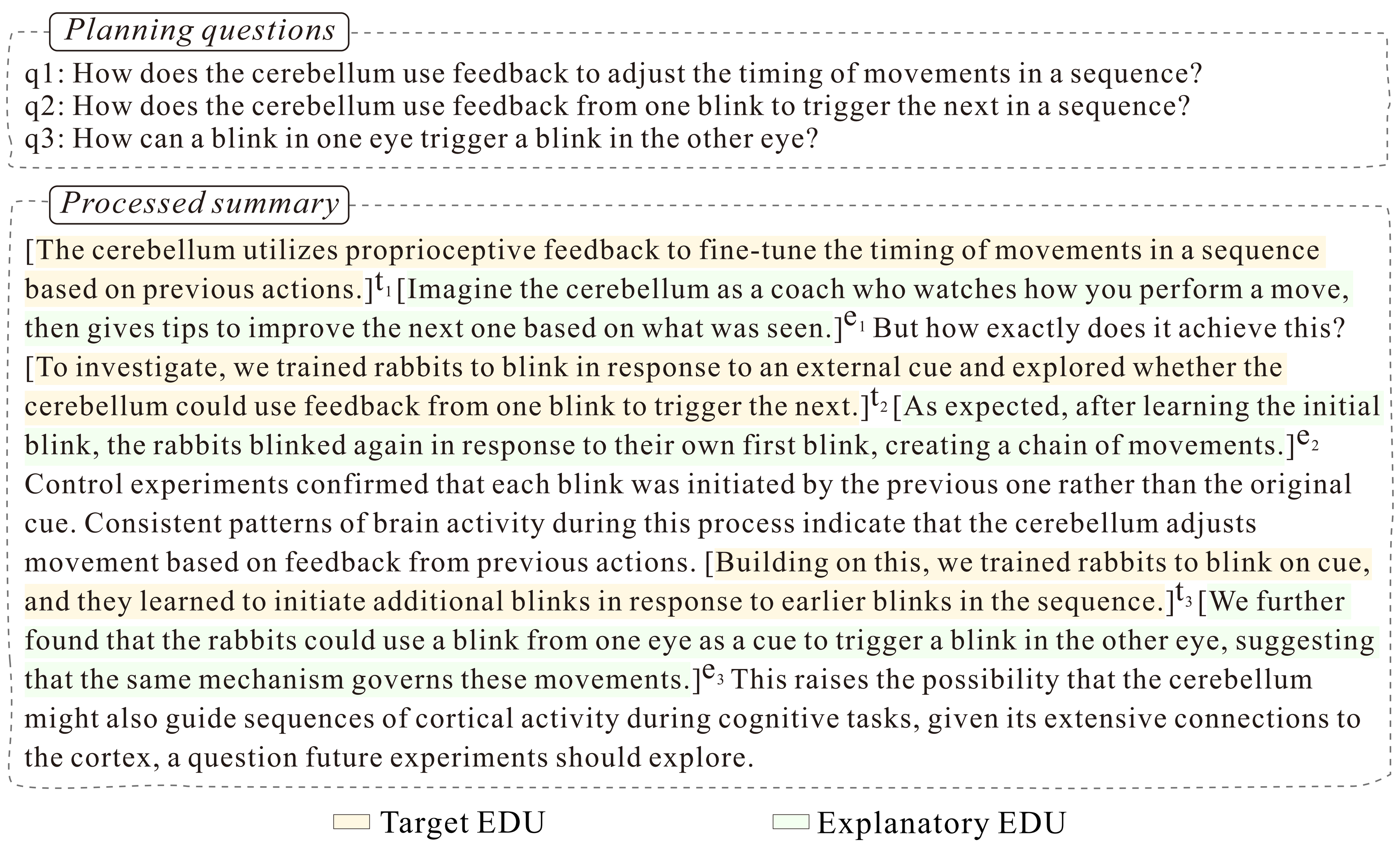

Figure 2. Plan generation pipeline

Method

Step 3: Two Model Variants

- **`Plan-Output`** (End-to-End): Train $P(b, s|x)$ to generate plan (b) + summary (s) jointly based on input document (x)
- **`Plan-Input`** (Pipeline): PG Module $P(b|x)$ generates plan, SG Module $P(s|x, b)$ generates summary

Main Results

- **`Plan-Input`** achieves **SOTA performance** across all metrics
- **More explanations**: ExpRatio 17.68% vs. 13.61% (`Mistral`$_{FT}$)
- **Better factual consistency**: VeriScore 0.71 vs. 0.56 (`Mistral`$_{FT}$)
- **Higher readability**: D-SARI 37.18 vs. 30.11 (`Mistral`$_{FT}$)

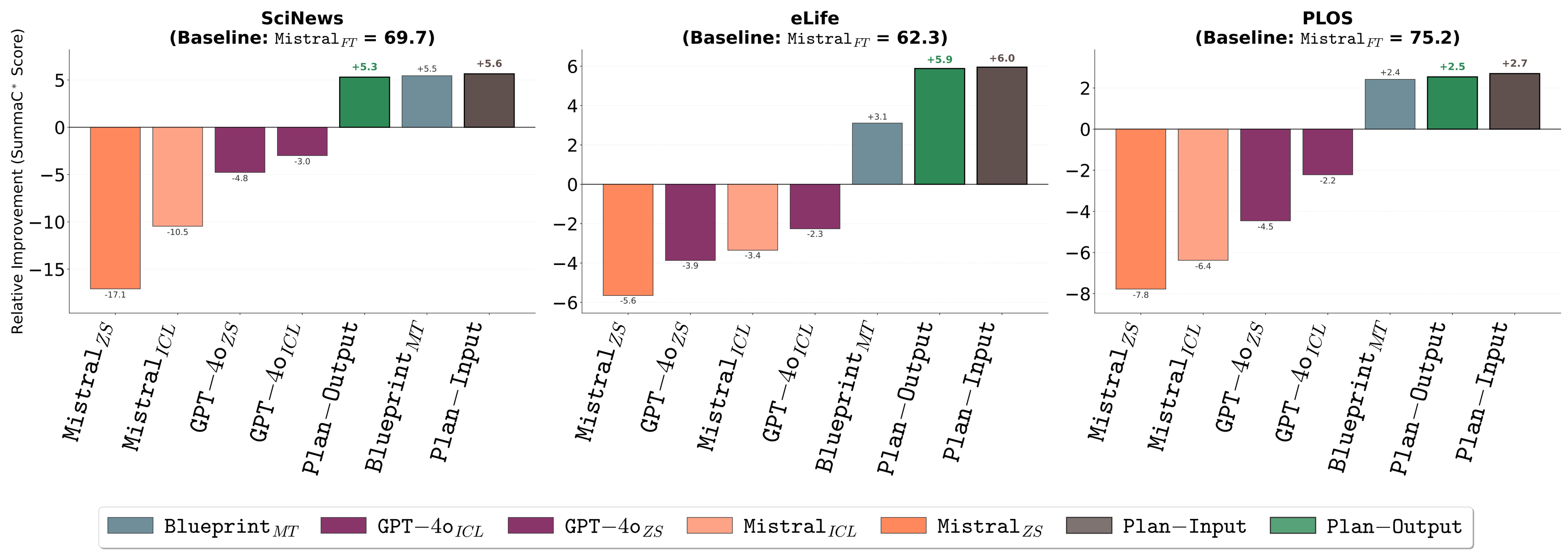

Figure 3. Model performance comparison across three datasets

Human Evaluation: Humans remain superior, but `Plan-Input` outperforms all baselines, achieving near human-level **accessibility** and substantial improvements in **explanation quality**.

Controllability of Explanatory Content

- Our models can **control explanation types** by modifying plans
- Deleting specific RST relation questions reduces the corresponding explanations in the output

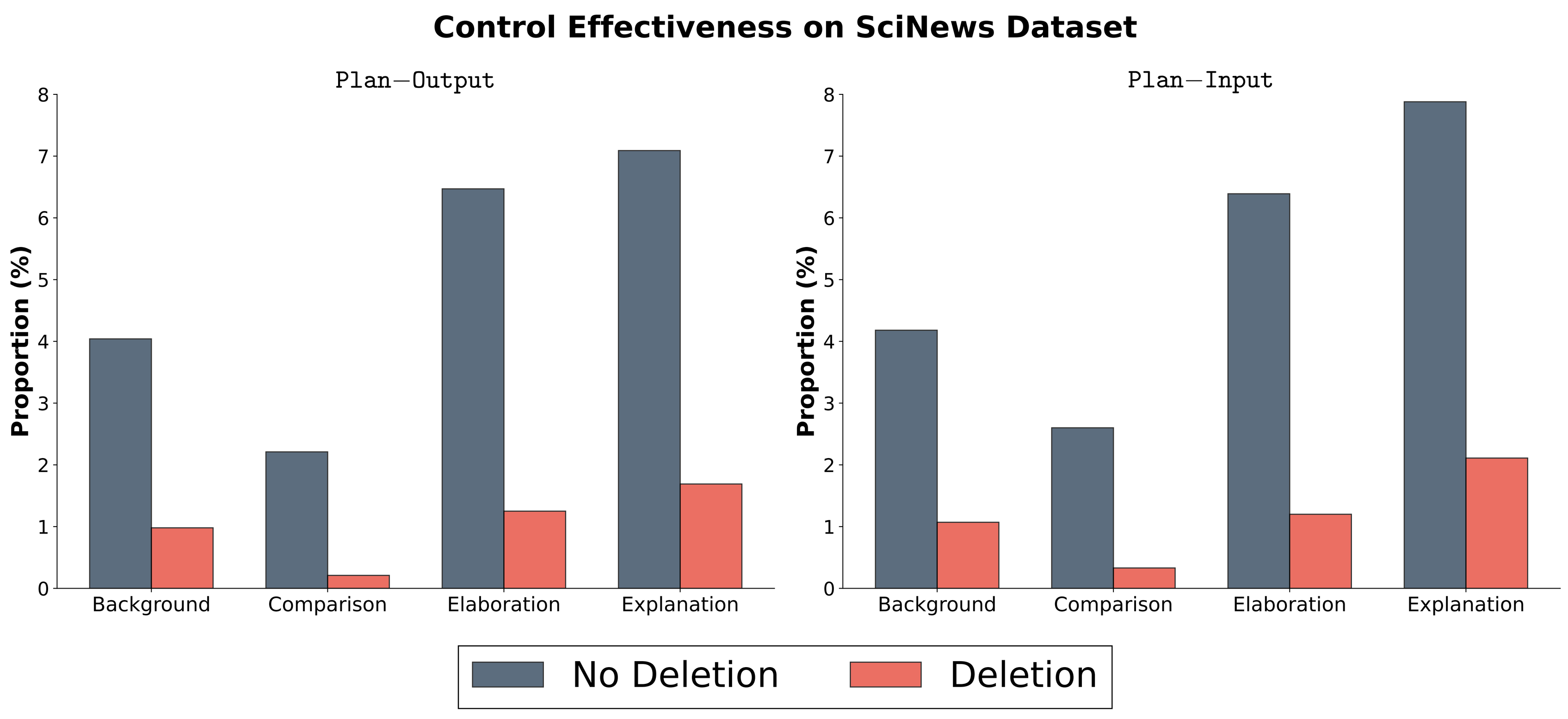

Figure 4. Control effectiveness across different explanation types

Conclusion

- Introduced **explanatory summarization** task for controlled lay summary generation
- Developed **discourse-driven planning** using RST + QUD frameworks
- Proposed two model variants: `Plan-Input` and `Plan-Output`
- Achieved **SOTA performance** on 3 datasets with significant improvements

Project Info

`dongqi.me/projects/ExpSum`

https://dongqi.me

Funded by the European Research Council (Grant No. 948878)

dongqi.me@gmail.com