

STRONG – Structure ConTRollable Legal OpiNion Summary Generation

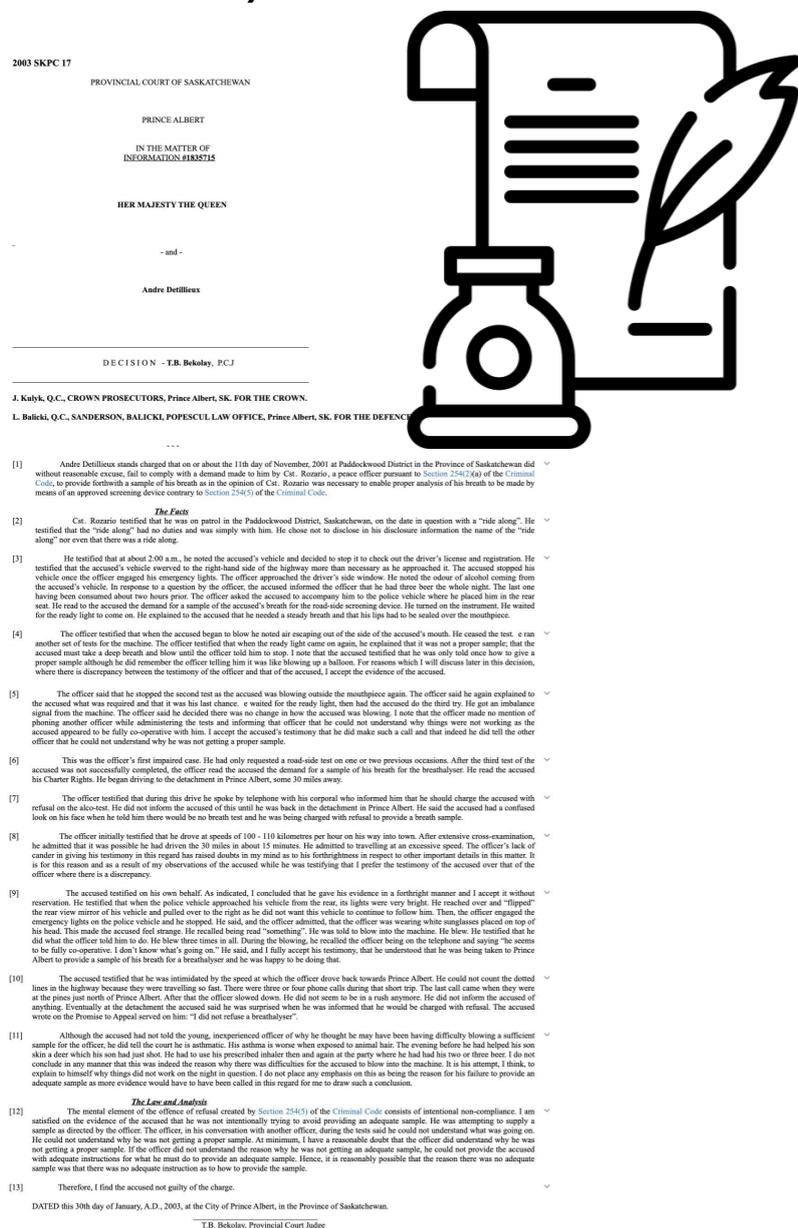
Yang Zhong and Diane Litman



Legal Case Summarization

Input: Legal case which is **long** and has diverse document structure

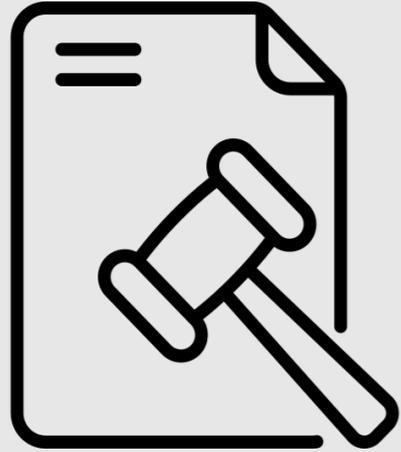
Output: A short paragraph composed by law community to summarize the argumentative components of the case (i.e. **Issue**, **Reasoning**, and **Conclusion**)*



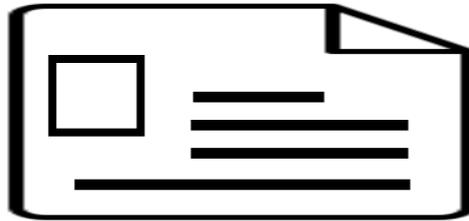
The accused was charged with failure to provide forthwith a breath sample contrary to Criminal Code s.254(5). HELD: The accused was found not guilty. The court accepted the evidence of the accused where there was a discrepancy between the testimony of the officer and the accused. The mental element of the offense of refusal consists of intentional non-compliance. The officer in his conversation with another officer said he could not understand why he was not getting a proper sample. There was a reasonable doubt that the officer did understand why he was not getting a proper sample and could not provide the accused with adequate instructions.

<https://www.CanLII.org/>

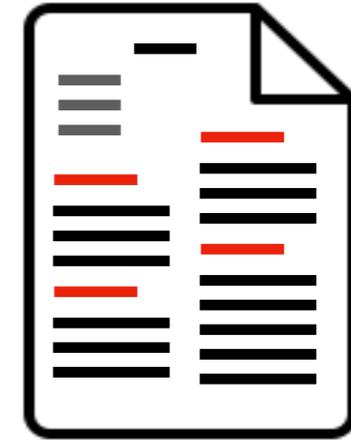
Challenges



CaseLaw Dataset
< 30k samples
Need Permission



News Dataset (CNN/DM)
> 300k samples



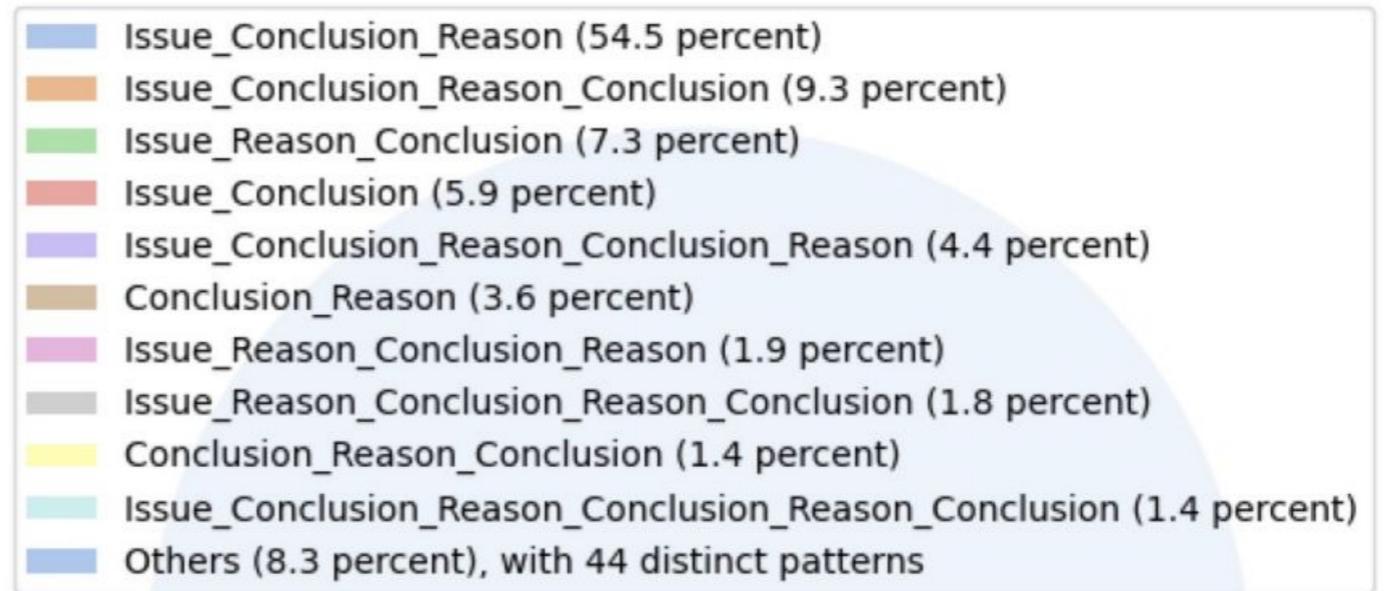
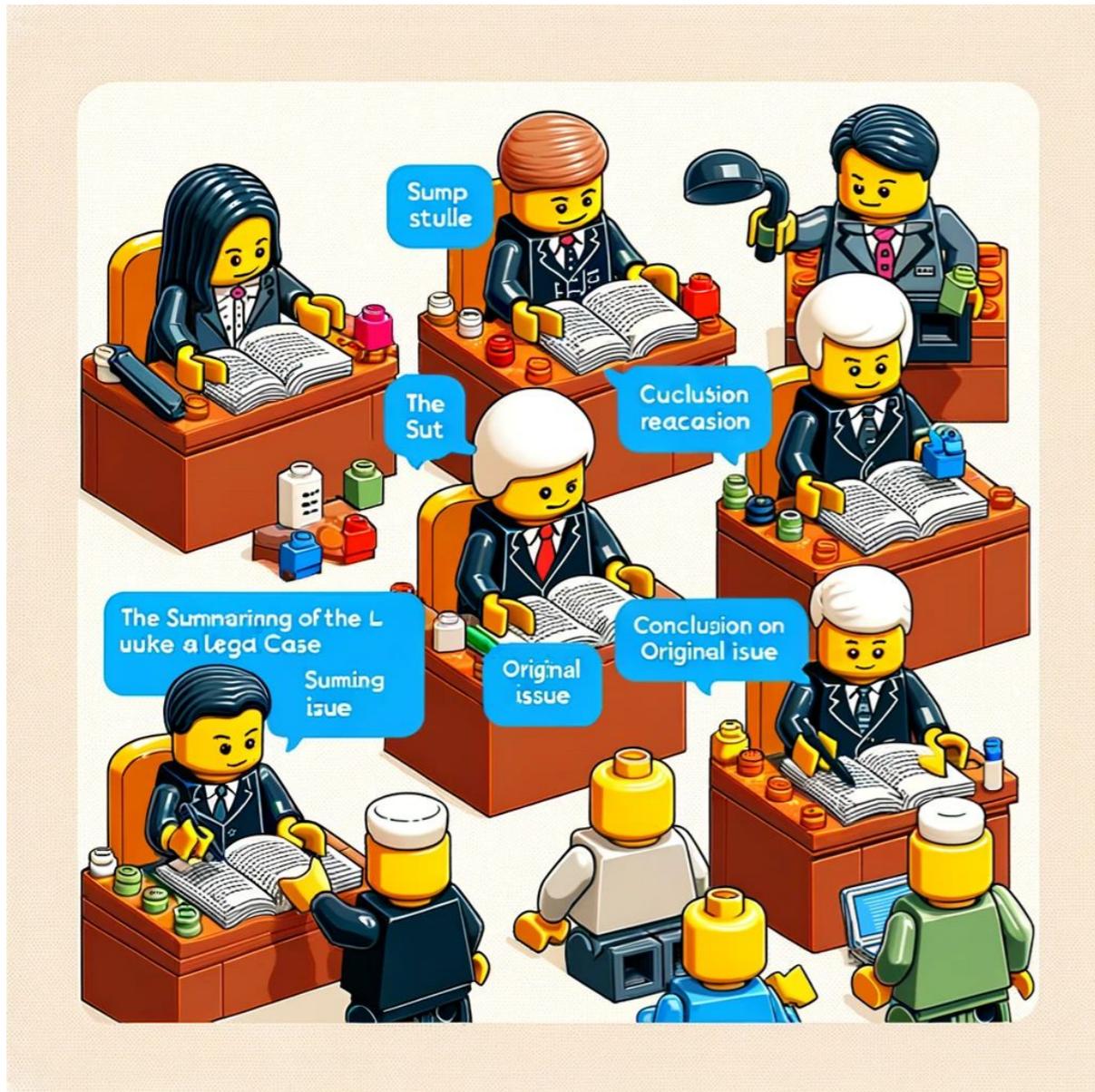
Scientific Dataset (arXiv)
> 200K samples

Legal Datasets are hard to access and of small scales

- ▶ The number of argumentative structural annotations is significantly lower
 - Only 1049 article-summary pairs are annotated*
 - The summaries are organized differently

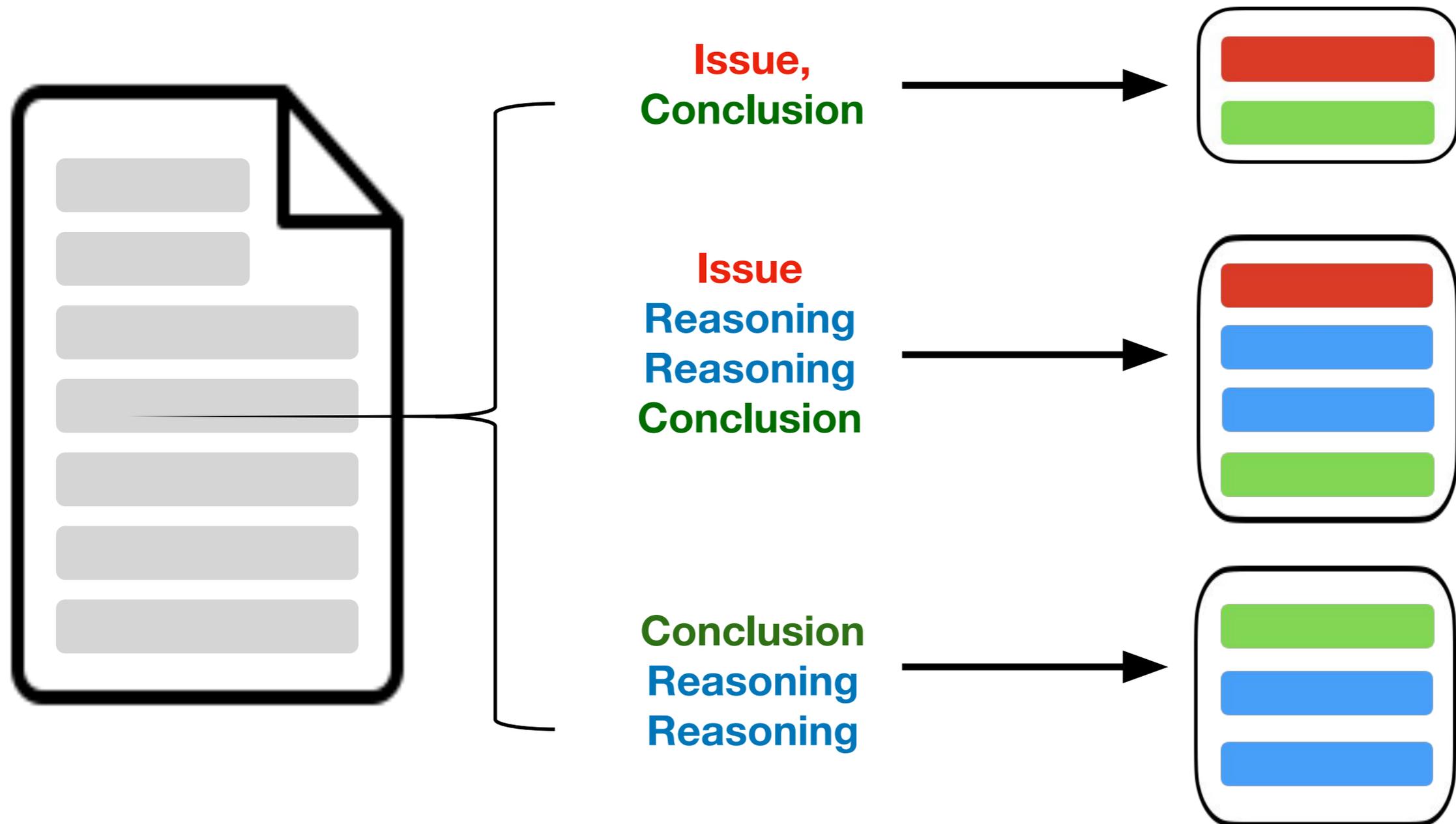
Summaries Can Be Organized Differently

The pattern **Issue**, **Conclusion**, followed by **Reasoning** sentences accounted for **54.5%** of all summarization patterns.



Controllable Abstract Summarization

Goal: Providing a structural plan, can we generate summaries accordingly?

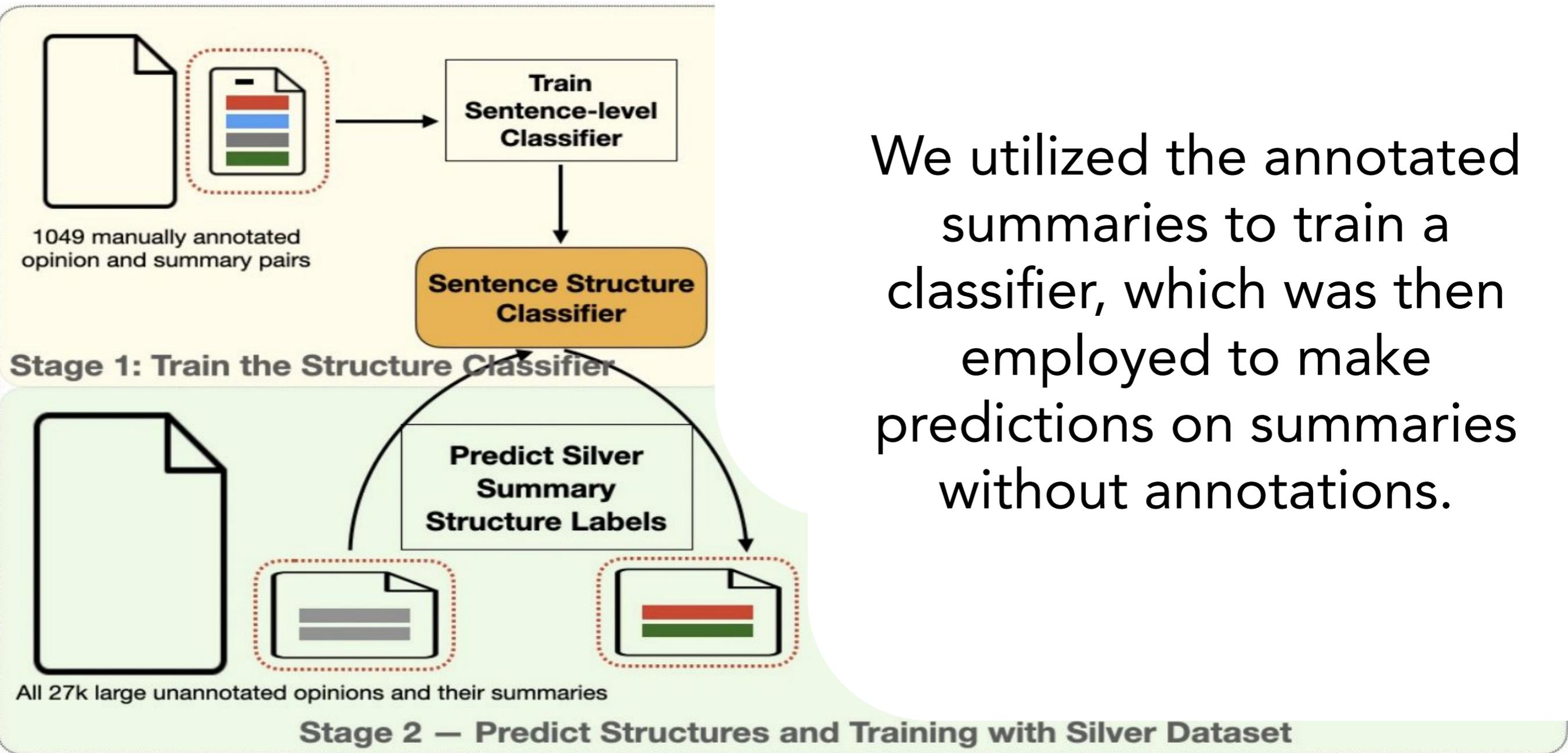


Our Work

- ▶ Augmentation of dataset to enrich the structure information
- ▶ A controllable generation framework to allow for summaries with different structures.

Our Work

- ▶ Augmentation of dataset to enrich the structure information
- ▶ A controllable generation framework to allow for summaries with different structures.



We utilized the annotated summaries to train a classifier, which was then employed to make predictions on summaries without annotations.

Utilize the Unlabeled Dataset

CanLII Dataset

- It consists of 28,290 legal opinions and human-written summary pairs. including 27,241 pairs with manually annotated labels.

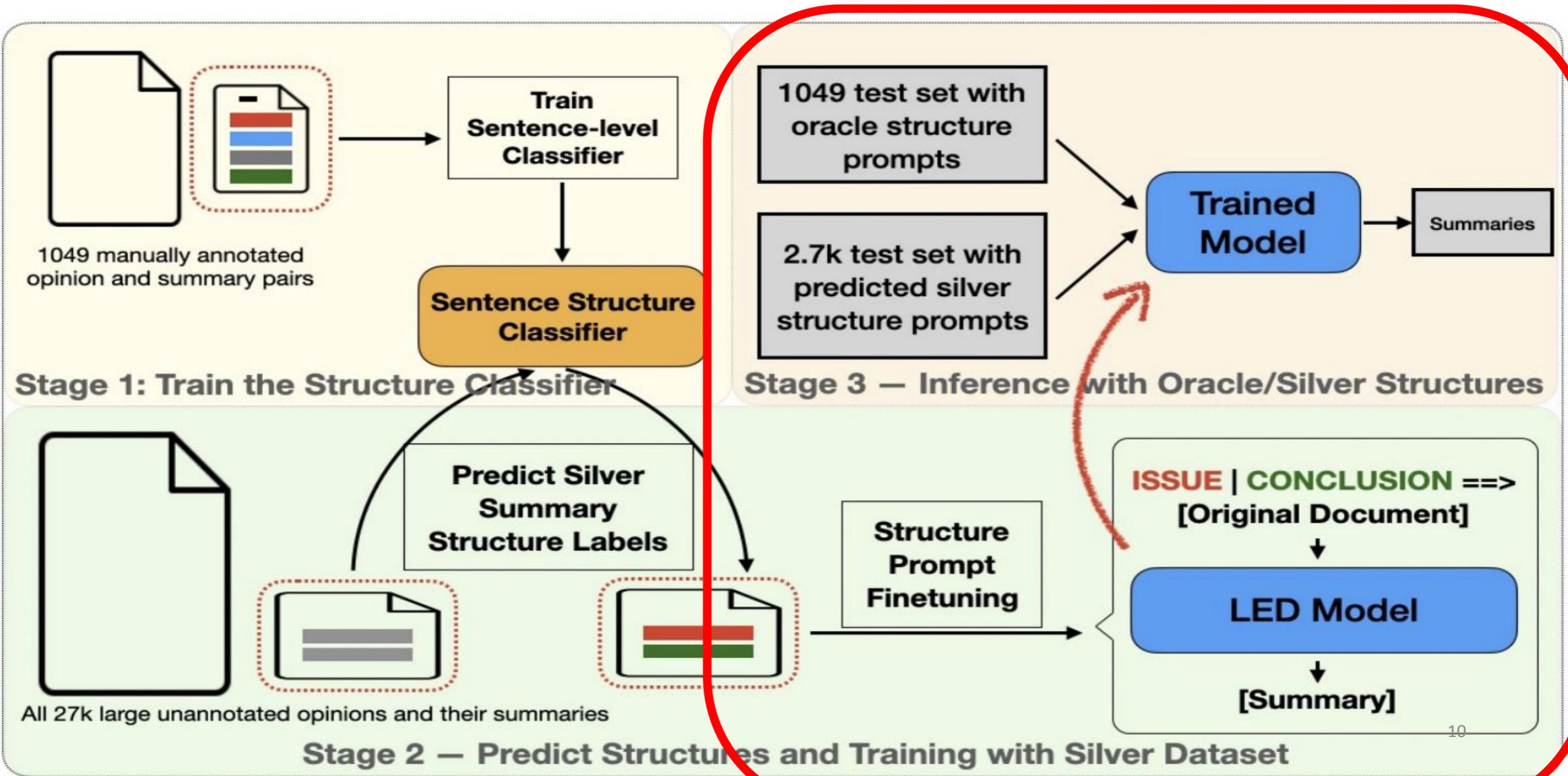
Our Work

- ▶ Augmentation of dataset to enrich the structure information
- ▶ A controllable generation framework to allow for summaries with different structures.

Structure Prompt Finetuning

Train the generation model by prompting the structure on silver dataset

Inference with oracle and silver structured data points



Dataset and Evaluation

- ▶ **Test Set:**

1049 case-summary pairs (with manual annotations)

2723 silver pairs with predicted structures

- ▶ **Metrics:** ROUGE-1/2/L F1 (Lin et al., 2004) and BERTScore (Zhang et al., 2020), Structure Similarity.

Dataset and Evaluation

▶ Test Set:

1049 case-summary pairs (with manual annotations)

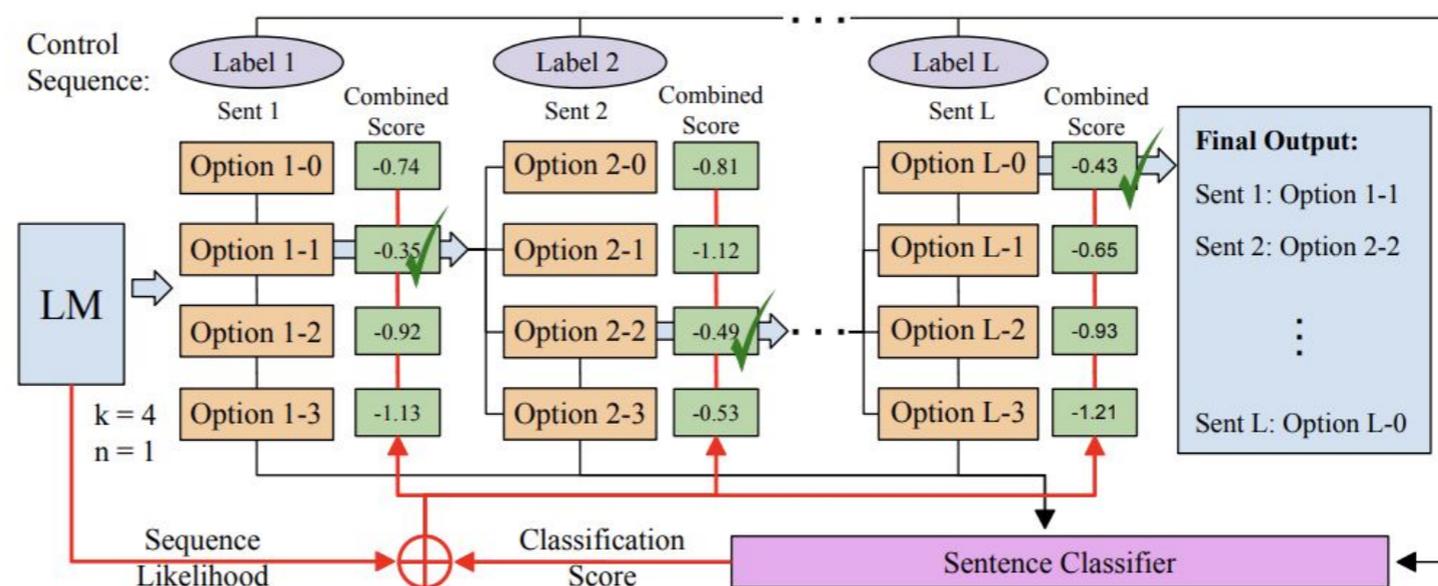
2723 silver pairs with predicted structures

▶ **Metrics:** ROUGE-1/2/L F1 (Lin et al., 2004) and BERTScore (Zhang et al., 2020), Structure Similarity.

▶ Baselines:

SentBS [1]

NoStructure



[1] SentBS: Sentence-level Beam Search for Controllable Summarization (Shen et al., EMNLP 2022)

Dataset and Evaluation

- ▶ **Test Set:**

1049 case-summary pairs (with manual annotations)

2723 silver pairs with predicted structures

- ▶ **Metrics:** ROUGE-1/2/L F1 (Lin et al., 2004) and BERTScore (Zhang et al., 2020), Structure Similarity

Reference
Summary



Generated
Summary



Dataset and Evaluation

- ▶ **Test Set:**

1049 case-summary pairs (with manual annotations)

2723 silver pairs with predicted structures

- ▶ **Metrics:** ROUGE-1/2/L F1 (Lin et al., 2004) and BERTScore (Zhang et al., 2020), Structure Similarity

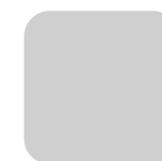
$$1 - \left(\frac{\text{minimum_edit_distance}(S_i, O_i)}{\max(\text{len}(S_i), \text{len}(O_i))} \right)$$

Reference
Summary



0.75

Generated
Summary



Experimental Result

- ▶ STRONG improves summarization quality compared to baselines
- ▶ STRONG generates summaries with higher structure similarities

	R-1	R-2	R-L	BS	
SentBS*	48.31	23.86	44.73	86.87	Max length 256
NoStructure	50.33	25.84	46.47	87.39	
STRONG	<u>52.47</u>	<u>26.54</u>	<u>48.57</u>	<u>87.63</u>	
NoStructure	51.61	26.72	47.76	78.49	Max length 512
STRONG	55.90	28.61	51.97	87.78	

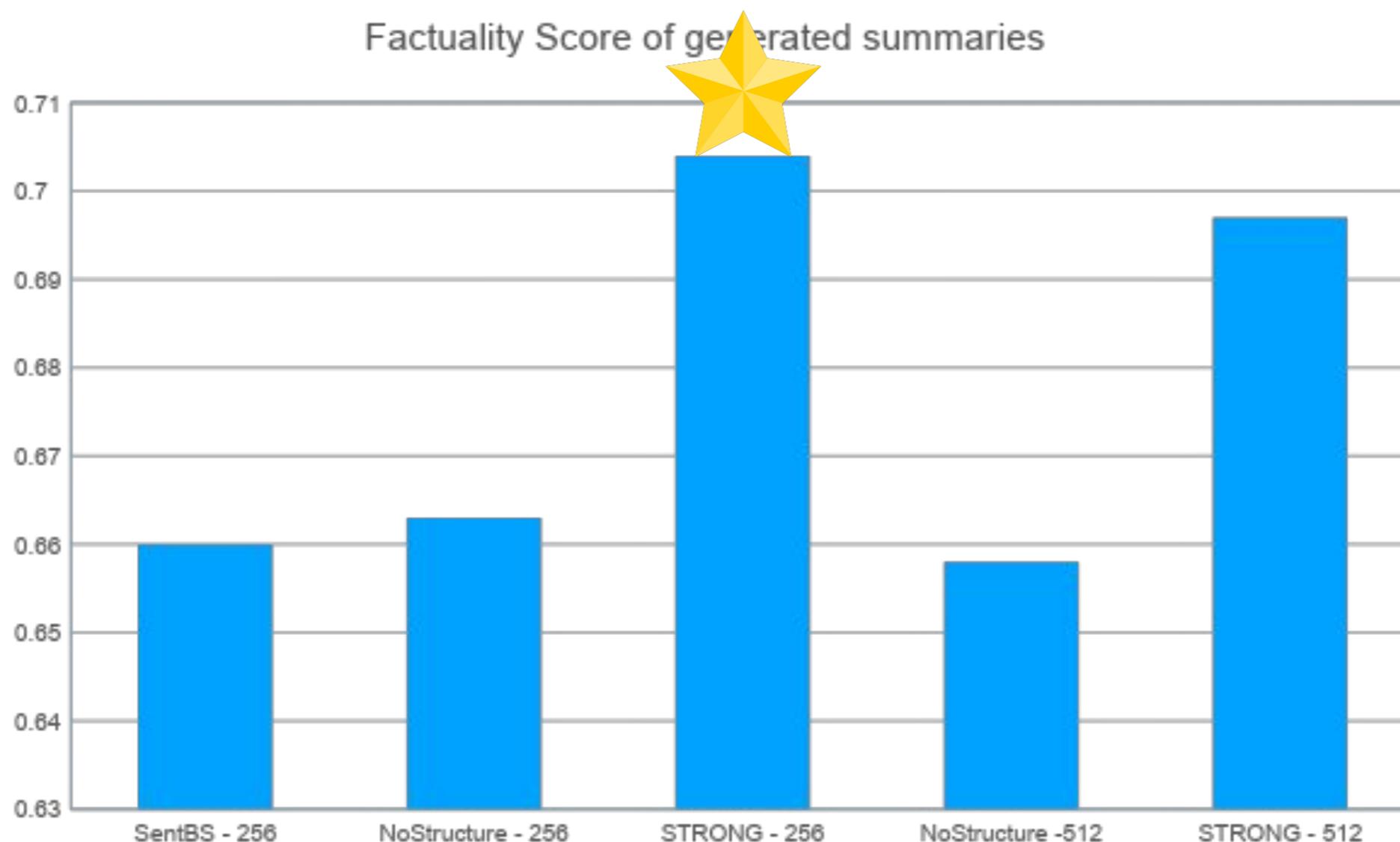
Experimental Result

- ▶ STRONG improves summarization quality compared to baselines
- ▶ STRONG generates summaries with higher structure similarities

	Structure Similarity	
SentBS*	0.436	Max length 256
NoStructure	0.344	
STRONG	0.372	
NoStructure	0.383	Max length 512
STRONG	0.535	

Factuality

- ▶ We picked the $\text{SUMMAC}_{\text{CONV}}$ score from Laban et al. (2022)
- ▶ Observation: structure information helps with highest consistency between document and generated summaries



Example (Controllability)

Prompt	Summary
<i>Human Reference</i>	
Issue Conclusion Conclusion Reason	<p>At issue was the interim custody of the child and child support. HELD: Interim custody was given to the petitioner with access to the respondent. The respondent was ordered to pay interim child support in the amount of \$193 per month. On an interim custody application, the Court must take into account the status quo.</p>
<i>STRONG generation results</i>	
Issue Conclusion Conclusion Reason	<p>At issue was the interim custody of the child and child support. The parties began cohabitating in May 1998 and separated in April 2000. HELD: The petitioner was awarded interim custody with reasonable access including every second weekend, reasonable telephone access and reasonable access during holiday periods. Both parents were ordered to complete the Parenting After Separation course. There was an unhealthy degree of constant bickering.</p>
Variant 1: Issue Conclusion Reason Reason Reason	<p>The petitioner was awarded interim custody with reasonable access including every second weekend, reasonable telephone access and reasonable access during holiday periods. Both parents were ordered to complete the Parenting After Separation course. The parties raised in their material the issue of each other's lack of cooperation and communication. While the Court was not in position to decide whether the accusations of each party were substantiated, it could say with relative certainty that there was an unhealthy degree of constant bickering. It was evident that as long as the parties work on establishing better communication, the child would benefit from contact with both parents with minimal disruption</p>

Takeaways

- ▶ STRONG improves summarization quality compared to baselines
- ▶ STRONG generates summaries with higher structure similarities

**STRONG – Structure ConTRollable Legal
OpiNion Summary Generation
Yang Zhong, Diane Litman**

Thank you!

Q & A

Backup

Length Control

Model	Control Len.	R-1	R-2	R-L	BS
NoStructure	No	50.33	25.84	46.47	87.39
STRONG	No	52.47	26.54	48.57	87.63
NoStructure	Yes	50.74	25.91	47.07	87.17
STRONG	Yes	50.96	26.26	47.33	87.39

Table 3: Results of models when summary has a maximum (top) versus controlled (bottom) length of 256 tokens. Although STRONG still outperforms the baseline, the delta is reduced when the length is controlled.