

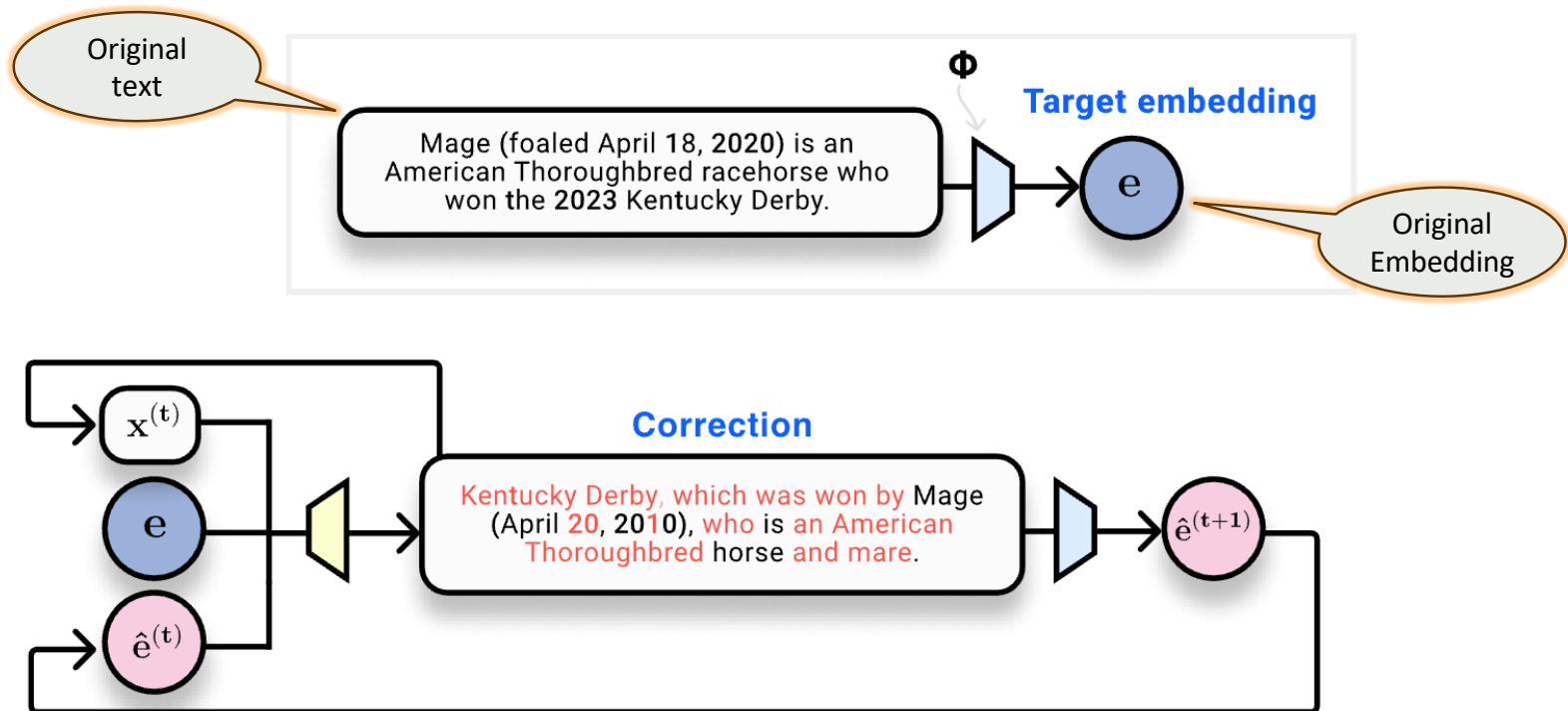


香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

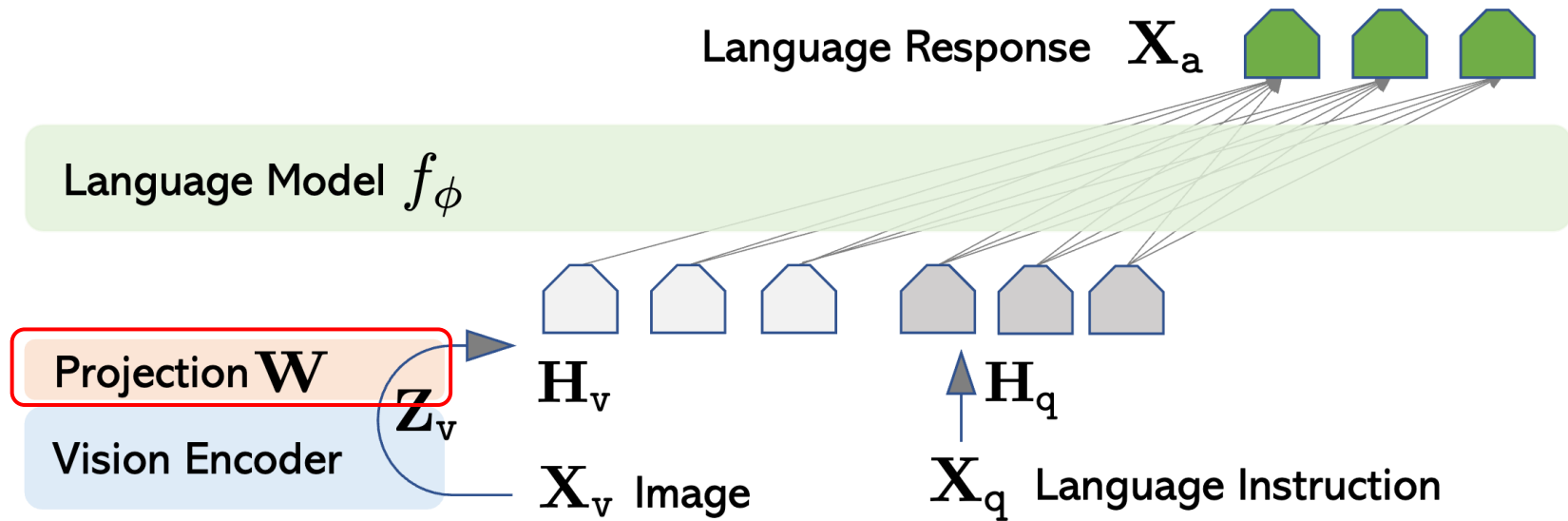
Rep2Text: Decoding Full Text from a Single LLM Token Representation

Haiyan Zhao, Zirui He, Yiming Tang, Fan Yang,
Ali Payani, Dianbo Liu, Mengnan Du

Relate Work – Embedding Inversion

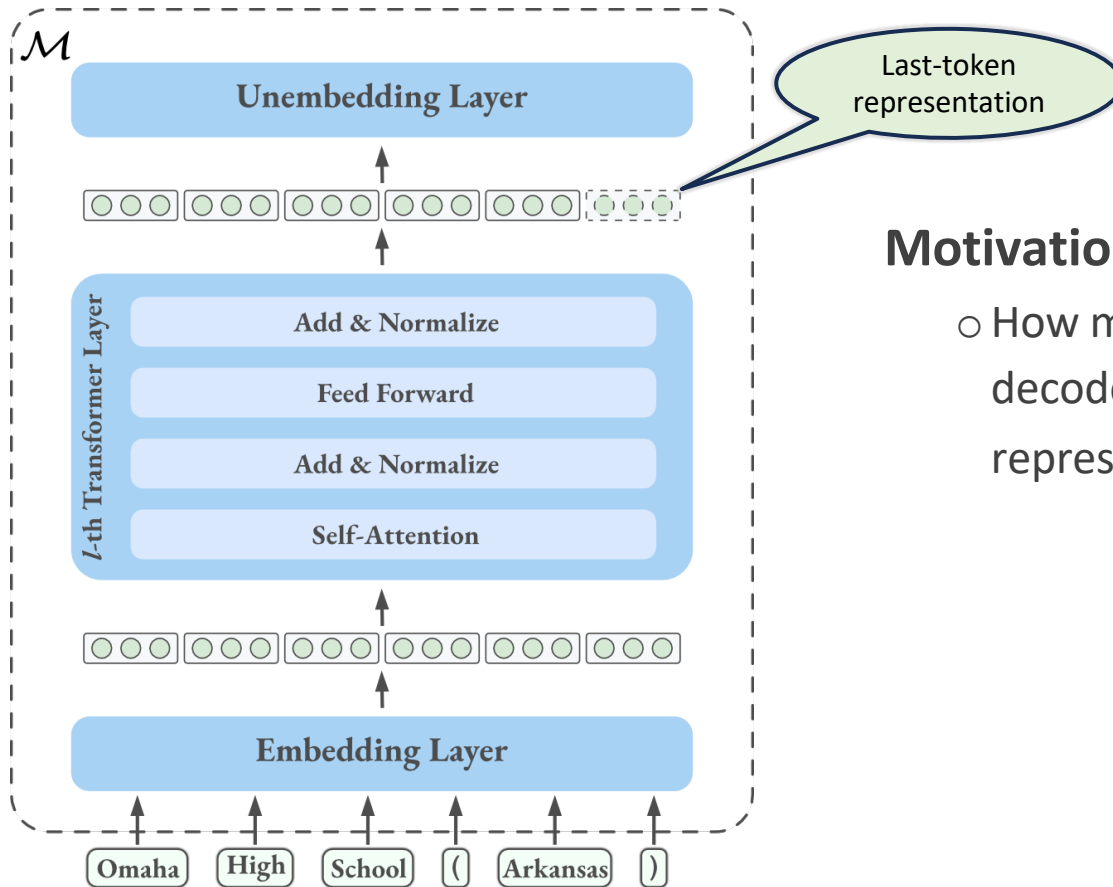


Relate Work – Vision Language Models



Similarly, can we align representation space with token embedding space?

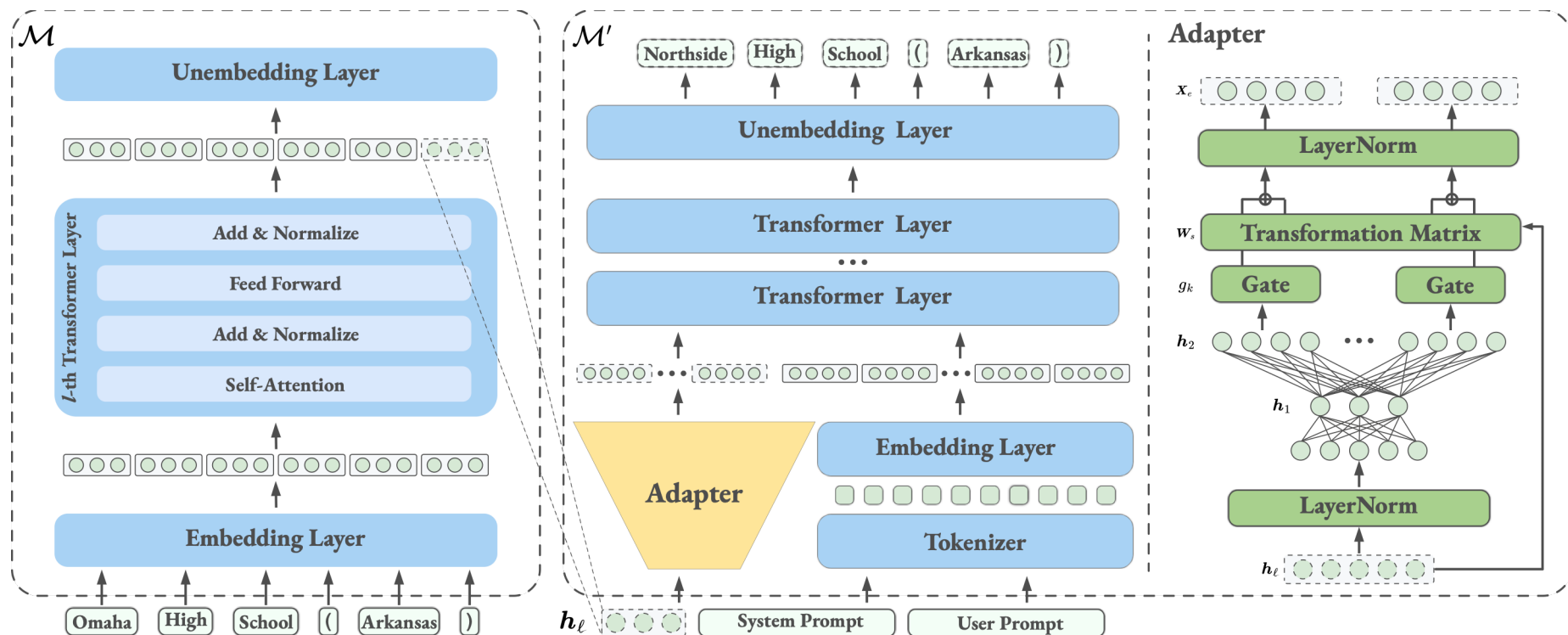
Our motivation



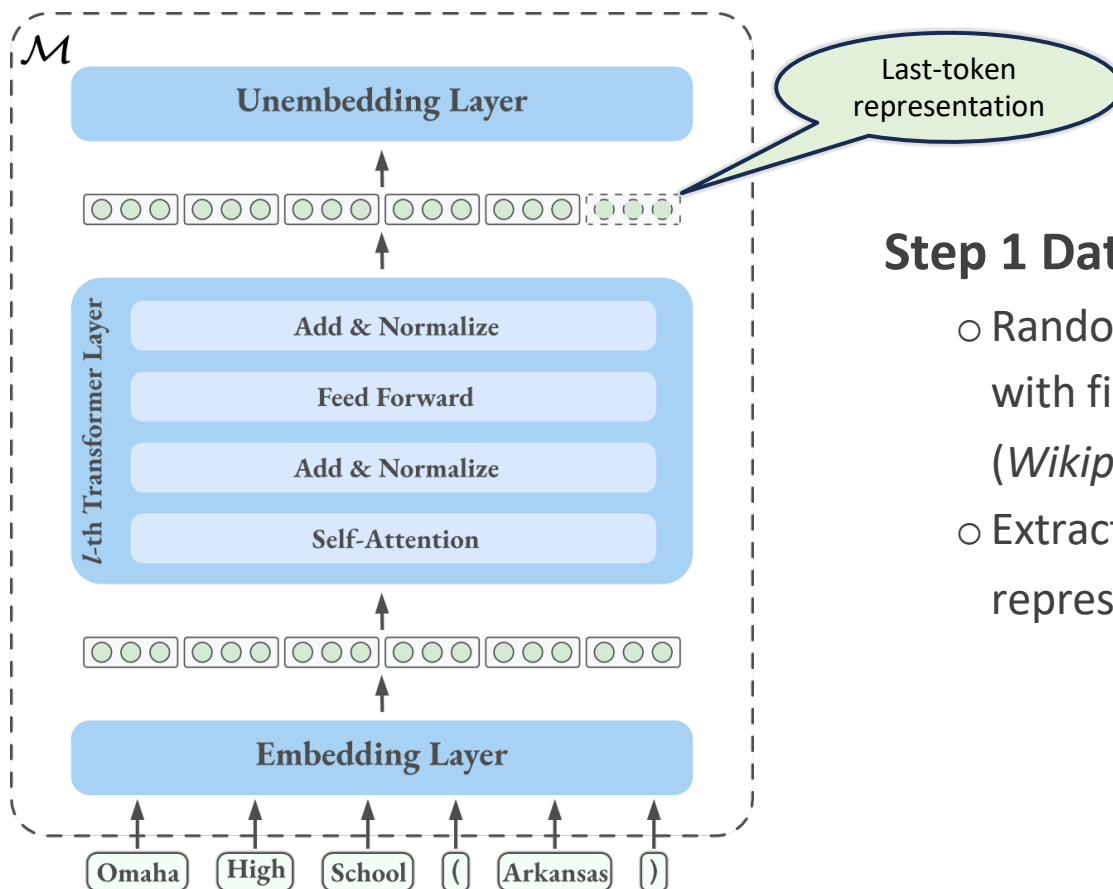
Motivation

- How much information can we decode from a single LLM token representation?

Our Method – Representation to Text (Rep2Text)



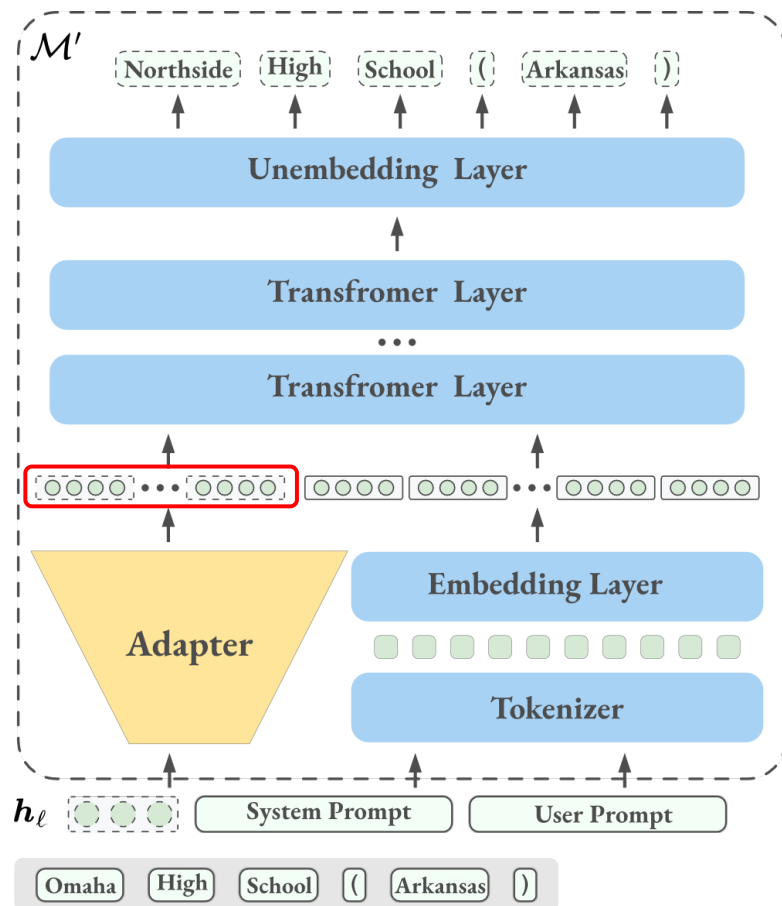
Our Method – Rep2Text



Step 1 Dataset Curation

- Randomly truncate input sequences with fixed number of tokens (*Wikipedia*)
- Extract the last token representation for each sequence

Our Method – Rep2Text



Step 2 Adapter training

- Map last token representation into the token embedding space
- Concatenation: <mapped token embedding, system prompt token embedding, user prompt token embedding>
- Passing through model layers and decode greedily

Evaluation of Rep2Text

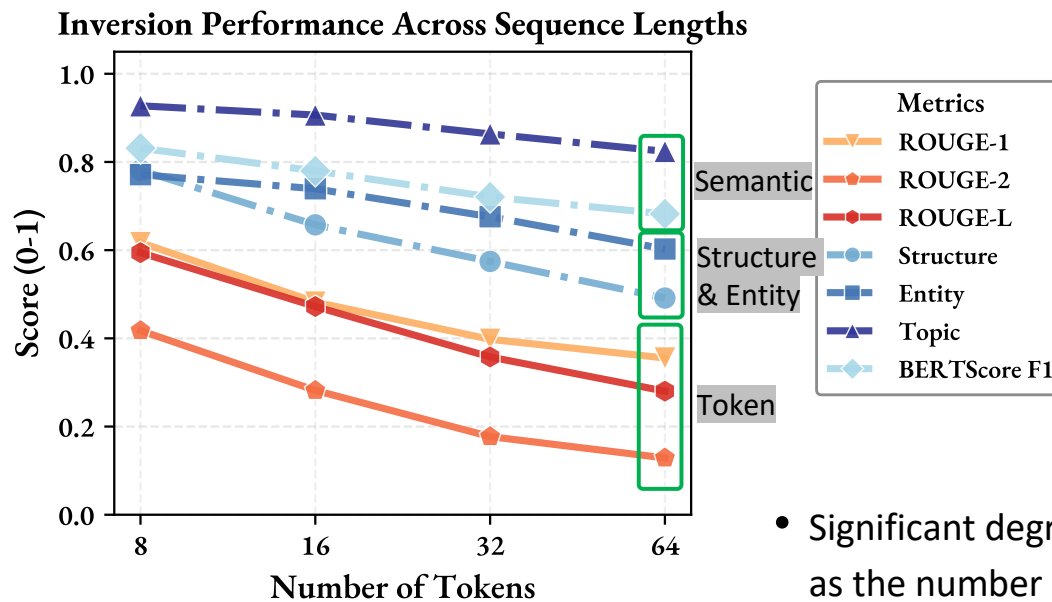
- Token-level Accuracy
 - ROUGE-1, ROUGE-2, ROUGE-L
- Sentence Structure and Entity Preservation
- Semantic Similarity

Ground-truth	[GT]	Dinwiddie Street Methodist church,	is a historic Methodist church located in Portsmouth
Generated	[GEN]	st. John's Episcopal church and Parish House	is a historic church located in Portsmouth
<hr/>			
	[GT]	Dinwiddie Street Methodist church,	is a historic Methodist church located in Portsmouth
	[GEN]	st. John's Episcopal church and Parish House	is a historic church located in Portsmouth
<div><div></div> Strong Structure Match <div></div> Weak Structure Match <div></div> Strong Entity Match <div></div> Weak Entity Match</div>			

Inversion Performance Regarding Number of Tokens

Tokens (#)	Ground-truth sequence	Inverted Sequence	ROUGE-1
8	in the Centre-Val de Loire	in the Centre-Val de Loire	1
16	species within the genus Conus, these snails are predatory and venomous.	species within the genus Conus, these snails are predatory and venomous.	1
32	1825 in Wales This article is about the particular significance of the year 1825 to Wales and its people.	1840 in Wales This article is about the particular significance of the year 1840 to Wales and its people.	0.91
64	Rush Hour 2 Rush Hour 2 is a 2001 American action comedy film directed by Brett Ratner and written by Jeff Nathanson, based on the characters created by Ross LaManna. A sequel to Rush Hour, it is the second installment in the Rush Hour series and stars Jackie Chan,	Rush Hour 2 Rush Hour 2 is a 2001 American buddy cop action comedy film directed by Brett Ratner and written by Jeff Nathanson. It is the sequel to the 1998 film Rush Hour and the second installment in the Rush Hour film series. The film stars Jackie Chan,	0.84

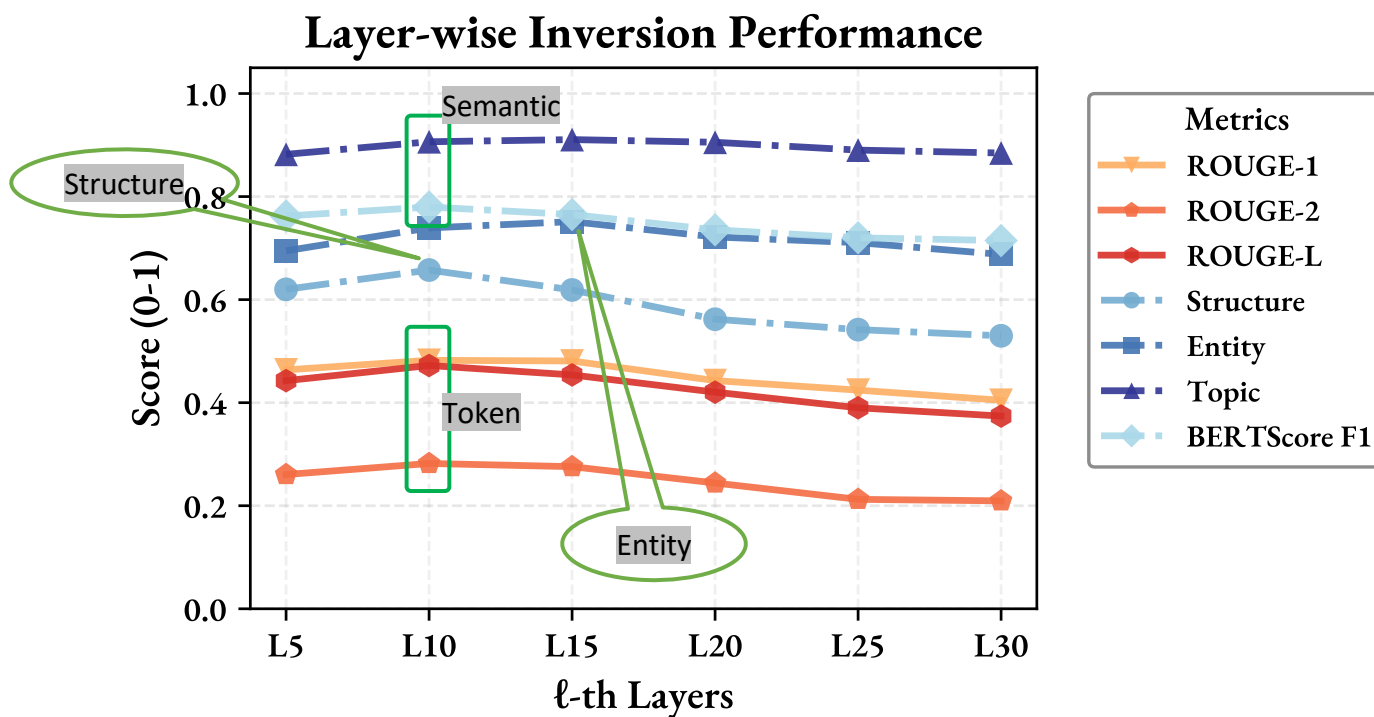
Inversion Performance Regarding Number of Tokens



- Structure and Entity is sensitive to number of tokens, too.
- Semantic-level information can always be preserved very well

- Significant degradation at the **token-level** inversion as the number of tokens increases
- More than half tokens can be exactly inverted when the sequence has less than 16 tokens

Inversion Performance Regarding ℓ -th Layer



- All metrics across layers are robust
- Slight improvement can be observed at Layer 10 and Layer 15

Generability of Rep2Text on Clinical Data

Ground-truth Sentences	Inverted Sequences	ROUGE-1	Structure Score	Entity Score	BERTScore	Topic Score
A. Rogers, 71, admitted for pneumonia, stable vitals.	://\n A. Rogers, 71, admitted for pneumonia, stable vitals. ://\n A. Rogers,	0.89	0.4	1	0.85	1
Karen J., 41, migraine recurrent, prescribed sumatriptan.	://www.medscape.com/viewarticle/509975\n Kaplan, M., 41, migraine recurrent	0.35	0.2	0.4	0.55	0.6
Pt denies pain, tolerating diet.	of 2.5 kg. 2.5 kg. 2.5 kg. 2.5 kg	0	0	0	0.35	0

Clinical data is fully OOD, yet many sentences invert well, **showing true generalization, not overfitting.**