

# Dynamics of Spontaneous Topic Changes in Next Token Prediction with Self-Attention

Mumin Jia

Joint work with  
Jairo Diaz-Rodriguez

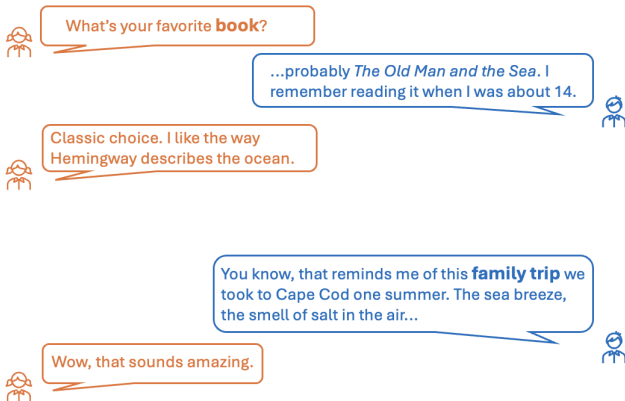
York University

Nov 6, 2025

# Table of Contents

- Motivation and Problem Statement
- Methodology
- Simulations
- Conclusion

# Spontaneous Human Thought



**Spontaneous topic change in human thought**



# Spontaneous Topic Changes in LLMs

Large Language Models (LLMs) follow a structured, statistical approach to generate answers.

# Spontaneous Topic Changes in LLMs

Large Language Models (LLMs) follow a structured, statistical approach to generate answers.

## Input:

My favorite book is *The Old Man and the Sea*.

## Chat GPT's Answer:

My favorite book is *The Old Man and the Sea*. I admire the quiet strength of Santiago, the old fisherman, and how he endures hardship with grace and dignity. The story's simplicity holds deep meaning—about perseverance, pride, and the relationship between humans and nature.



LLMs keep input topic



# Spontaneous Topic Changes in LLMs vs. in Human Thought

- Suppose a Large Language Model (LLM) gains knowledge of **Book**.

# Spontaneous Topic Changes in LLMs vs. in Human Thought

- Suppose a Large Language Model (LLM) gains knowledge of **Book**.



# Spontaneous Topic Changes in LLMs vs. in Human Thought

- Suppose a Large Language Model (LLM) gains knowledge of **Book**.



- Suppose a Large Language Model (LLM) gains knowledge of both **Book and Trip**.



# Spontaneous Topic Changes in LLMs vs. in Human Thought

- Suppose a Large Language Model (LLM) gains knowledge of **Book**.



- Suppose a Large Language Model (LLM) gains knowledge of both **Book** and **Trip**.

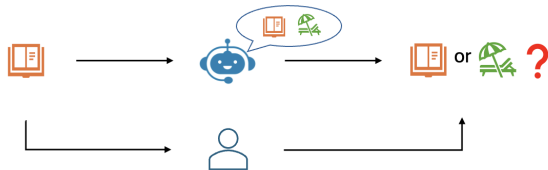


# Spontaneous Topic Changes in LLMs vs. in Human Thought

- Suppose a Large Language Model (LLM) gains knowledge of **Book**.



- Suppose a Large Language Model (LLM) gains knowledge of both **Book and Trip**.



# Problem Statement

What's the difference between **spontaneous topic changes** in **Large Language Models** and those in **human thought**?

# Table of Contents

- Motivation and Problem Statement
- Methodology
- Simulations
- Conclusion

# Defining Topics

- Token Priority Graph (TPG) (Li et al., 2024):
  - Tokens within the same dashed rectangular share equal priority.
  - Edges represent priority orders.

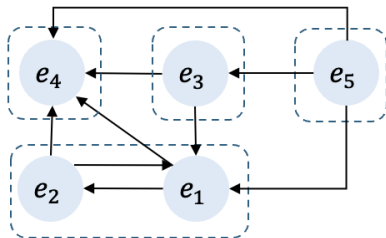


Figure 1: An example of Token Priority Graph (TPG).

# Defining Topics

- Token Priority Graph (TPG) (Li et al., 2024):
  - Tokens within the same dashed rectangular share equal priority.
  - Edges represent priority orders.
- We define a **Topic** as a set of **TPGs**.

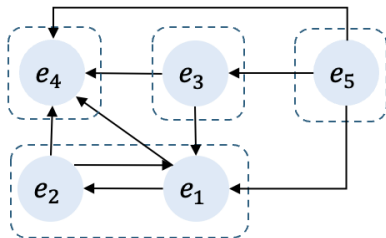


Figure 1: An example of Token Priority Graph (TPG).

# Table of Contents

- Motivation and Problem Statement
- Methodology
- **Simulations**
- Conclusion


# In Self-Attention Models

To validate Theorem 4, we use **synthetic data** under the **single-layer self-attention** model.

- Vocabulary size  $K = 10$ .
- 5000 input sequences under each parameter setting.



# In Self-Attention Models

Probability of Spontaneous Topic Changes	in Self-Attention Models
a. Input length	
b. Topic overlap	

# In Self-Attention Models

Probability of Spontaneous Topic Changes	in Self-Attention Models
a. Input length	
b. Topic overlap	

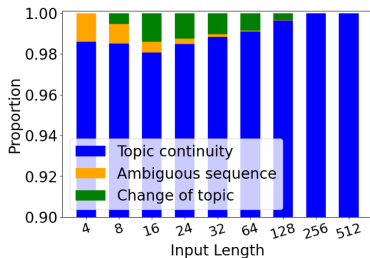


Figure 2: Impact of input length in self-attention model.

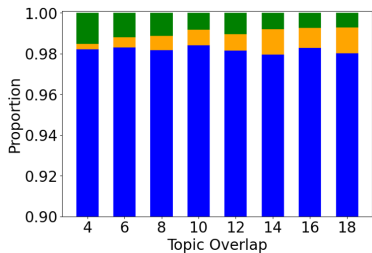


Figure 3: Impact of topic overlap in self-attention model.

# In Large Language Models (LLMs)

To further validate Theorem 4, we extend experiments in **frontier LLMs** with **real dataset**.

# In Large Language Models (LLMs)

To further validate Theorem 4, we extend experiments in **frontier LLMs** with **real dataset**.


- We measure the **cosine similarity** of two responses:
  - LLM has knowledge from paper A,
  - LLM has knowledge from both paper A and paper B.

# In Large Language Models (LLMs)

To further validate Theorem 4, we extend experiments in **frontier LLMs** with **real dataset**.

- We measure the **cosine similarity** of two responses:
  - LLM has knowledge from paper A,
  - LLM has knowledge from both paper A and paper B.
- **Larger** cosine similarity indicates **Lower** probability of spontaneous topic changes.

# In Large Language Models (LLMs)

Probability of Spontaneous Topic Changes	in Self-Attention Models
a. Input length	
b. Topic overlap	

# In Large Language Models (LLMs)

Probability of Spontaneous Topic Changes	in Self-Attention Models
a. Input length	
b. Topic overlap	

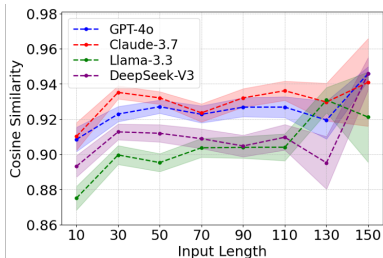


Figure 4: Impact of input length in LLMs.

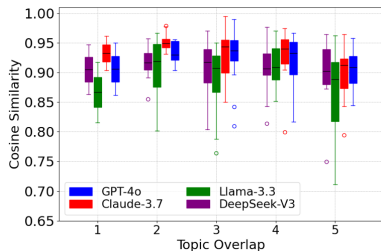


Figure 5: Impact of topic overlap in LLMs.

# Table of Contents

- Motivation and Problem Statement
- Methodology
- Simulations
- Conclusion



# Conclusion

- In human cognition, extended discussions and greater topic connectivity encourage spontaneous topic changes.
- In LLMs, neither longer inputs nor greater topic ambiguity increases the likelihood of a spontaneous topic change.