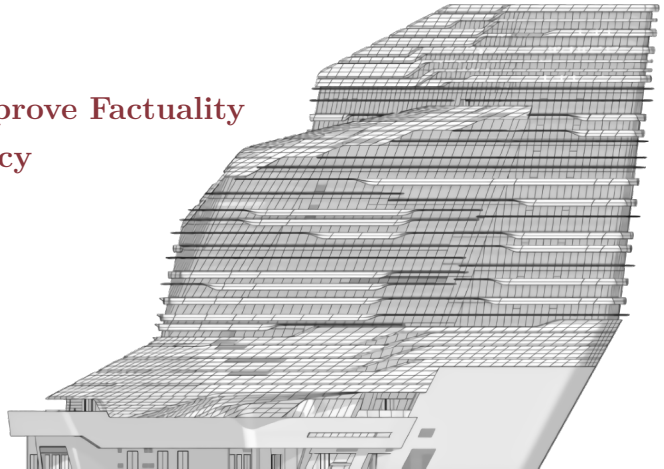
# Integrative Decoding: Improve Factuality via Implicit Self-consistency

Yi Cheng

The Hong Kong Polytechnic University



# The Challenge of Hallucinations

- LLMs often generate plausible but incorrect information.
- This phenomenon is known as "**hallucinations**."

# Self-consistency as a Hallucination Indicator

**Sampled Response #1**: David Parnas is credited with the concept of "information hiding", which emphasizes structuring software into manageable, self-contained modules... His work on software architecture principles, including the "Parnas Principle", which advocates for the separation of concerns, has had a lasting impact on the design and development of complex software systems.

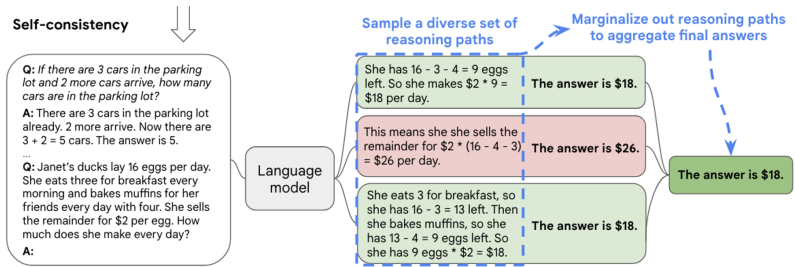
**Sampled Response #2**: David Parnas introduced the principle that "information hiding" is a key concept in software design, advocating that ... He contributed significantly to the concept of modular programming, emphasizing that ... Parnas was instrumental in the development of software architecture theory, focusing on the design of systems that are understandable, scalable, and adaptable ...

**Sampled Response #3**: David Parnas introduced the concept of "information hiding" in software engineering, ... Parnas was a pioneer in the field of software architecture, advocating for the importance of understanding and managing the structure and organization of complex software systems. He contributed to the development of the "Parnas Principle," which suggests that a program should be understandable by a single individual ...

**Sampled Response #4**: Parnas advocated for modular programming, emphasizing that ... Parnas contributed to the field of software architecture, emphasizing the importance of architecture in ... Parnas recognized the importance of human factors in software engineering, advocating for the integration of psychological and sociological aspects into the design and development processes of software systems.

- **Self-consistency (SC)** measures the consistency level among different outputs generated by LLMs in response to the same prompt. Statements consistently present across multiple sampled responses are more likely to be factual.
- SC serves as a useful indicator for hallucination detection and uncertainty quantification. It helps identify non-factual content.

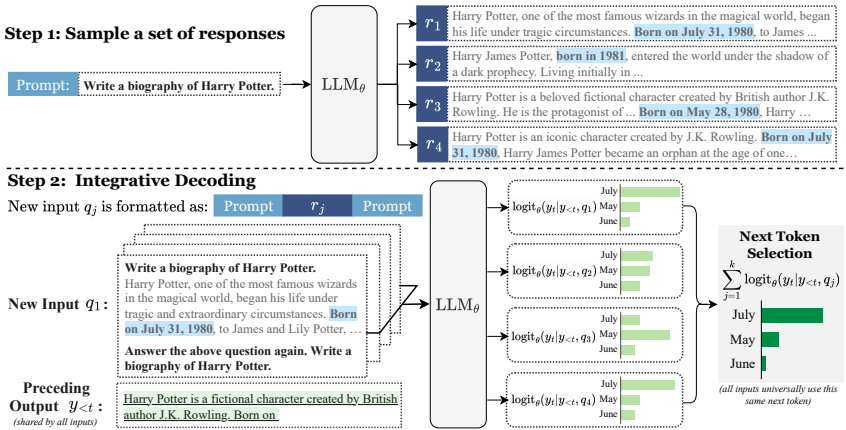
# Improve Factuality with Self-consistency




- Previous research leverages SC to mitigate hallucination, but **existing approaches pose strict constraints on task formats**, limiting their applicability to tasks with exact-match answers (e.g., arithmetic problems).
- **How to apply them to open-ended and long-form generation efficiently?**

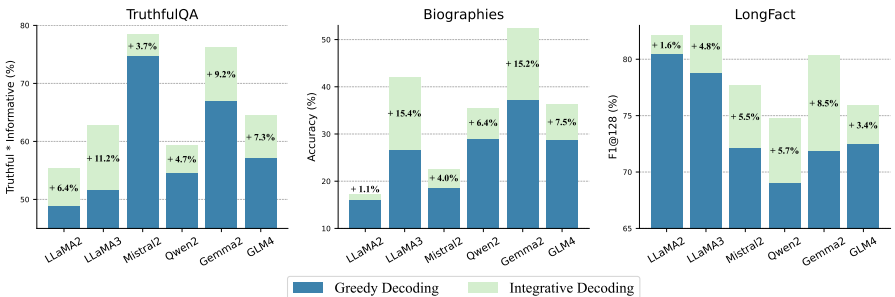
# Integrative Decoding (ID)

- ID implicitly incorporates self-consistency into the decoding objective.



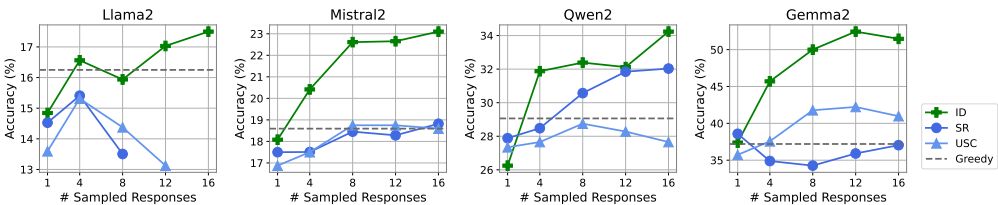
# Key Results: Improved Factuality

- With no need of retrieving external knowledge and additional training, integrative decod- ing consistently improves the factuality performance over six types of large language models, with substantial improvements on the TruthfulQA, Biographies, and LongFact datasets.



# Advantages of ID

- Balances factuality and informativeness.
- Robust to document-level generation tasks.
- Scales well with more sampled responses.



# Conclusion and Future Work

- We propose integrative decoding, a scalable, efficient, and effective decoding strategy that improves factuality in LLMs.
- Future research directions include combining ID with speculative decoding for further efficiency improvement and exploring more precise ways to approximate the self-consistency objective.

# Q&A

*Thank you for listening!*