# Robust Information Retrieval

SIGIR 2024 tutorial

**Yu-An Liu**[a,b], Ruqing Zhang[a,b], Jiafeng Guo[a,b] and **Maarten de Rijke**[c]

https://sigir2024-robust-information-retrieval.github.io/

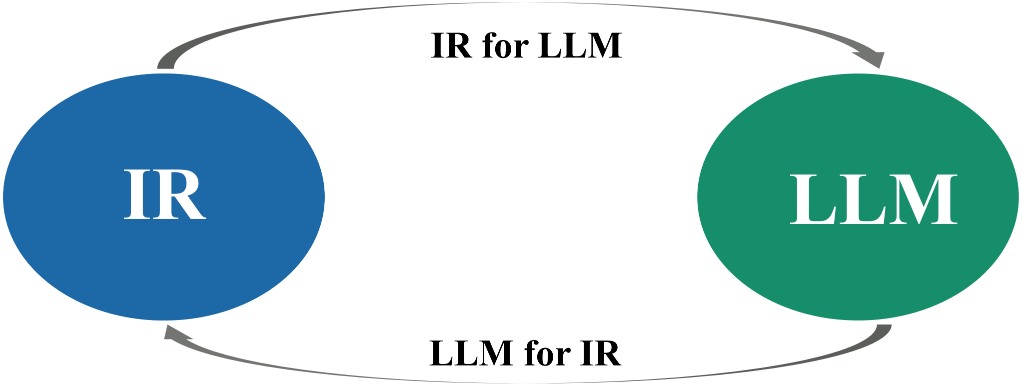
July 14, 2024

01:30 – 05:00 PM

[a] Institute of Computing Technology, Chinese Academy of Sciences
[b] University of Chinese Academy of Sciences
[c] University of Amsterdam

# Section 5:
# Robust IR in the age of LLMs

# IR in the age of LLMs




- **IR for LLM**: Retrieval-augmented generation
- **LLM for IR**: A double-edged sword

# New opportunities for IR robustness via LLMs

# New opportunities to adversarial robustness

LLMs hold promise for improving the adversarial robustness of IR systems through their ability to generate and identify adversarial examples:

# New opportunities to adversarial robustness

LLMs hold promise for improving the adversarial robustness of IR systems through their ability to generate and identify adversarial examples:

- **Generating adversarial examples with LLMs**
  - AIGC scenario
  - Superior capabilities in language generation and interaction
  - Hardening the IR system with generated adversarial samples

# New opportunities to adversarial robustness

LLMs hold promise for improving the adversarial robustness of IR systems through their ability to generate and identify adversarial examples:

- **Generating adversarial examples with LLMs**
  - AIGC scenario
  - Superior capabilities in language generation and interaction
  - Hardening the IR system with generated adversarial samples
- **Adversarial defense assisted with LLMs**
  - Identifying adversarial samples
  - Enhancing existing defense strategies

# New opportunities to OOD robustness

The powerful generation and language understanding capability of LLMs can help to improve the OOD robustness of IR systems:

# New opportunities to OOD robustness

The powerful generation and language understanding capability of LLMs can help to improve the OOD robustness of IR systems:

- **Synthesizing OOD training data with LLMs**
  - LLMs can generate diverse and complex datasets that mirror OOD scenarios
  - Synthetic data can help improve the generalizability and robustness of IR models against OOD inputs

# New opportunities to OOD robustness

The powerful generation and language understanding capability of LLMs can help to improve the OOD robustness of IR systems:

- **Synthesizing OOD training data with LLMs**
  - LLMs can generate diverse and complex datasets that mirror OOD scenarios
  - Synthetic data can help improve the generalizability and robustness of IR models against OOD inputs
- **LLMs for OOD detection**
  - With capabilities of language understanding, LLMs can detect OOD queries
  - Neural IR models may perform worse on these OOD queries that deviate from the training distribution

# New challenges for IR robustness via LLMs

# New challenges to adversarial robustness

When applied to IR systems, the adversarial vulnerability of the LLMs themselves is introduced at the same time, as demonstrated by:

# New challenges to adversarial robustness

When applied to IR systems, the adversarial vulnerability of the LLMs themselves is introduced at the same time, as demonstrated by:

- **The vulnerability caused by hallucinations of LLMs**
- **Defense costs associated with the scale and opacity of LLMs**

# New challenges to adversarial robustness

The vulnerability caused by hallucinations of LLMs

- With hallucination, LLMs can generate plausible yet factually incorrect information
- Such reliance can undermine the trustworthiness and reliability of the IR system

**Prompt:** Please **rank** the following documents according to their relevance to the query {{query}} and output the document IDs.
[1]{Doc_1}, [2]{Doc_2}, ..., [n]{Doc_n}

Sure! I can help you. The relevance ranking is:
[2] > [3] > 358 > [1] ##3 > 68235 > ....

# New challenges to adversarial robustness

Defense costs associated with the scale and opacity of LLMs

- LLMs operate as black boxes with limited transparency into how decisions are made
- This opacity complicates efforts to diagnose and mitigate vulnerabilities

# New challenges to OOD robustness

LLMs have shown biases and input sensitivities in existing work, and these will affect the OOD robustness of IR systems:

# New challenges to OOD robustness

LLMs have shown biases and input sensitivities in existing work, and these will affect the OOD robustness of IR systems:

- **Bias in the corpus domain of LLMs**
  - The training process of LLMs leads to a bias towards the domain characteristics
  - This can degrade performance when the model encounters OOD queries or documents

# New challenges to OOD robustness

LLMs have shown biases and input sensitivities in existing work, and these will affect the OOD robustness of IR systems:

- **Bias in the corpus domain of LLMs**
  - The training process of LLMs leads to a bias towards the domain characteristics
  - This can degrade performance when the model encounters OOD queries or documents
- **Sensitivity of LLMs to query inputs**
  - LLMs can exhibit high sensitivity to slight variations in input
  - This potentially leads to significantly different IR outcomes

# So much to do ...

**Making robustness one of the hallmarks of IR in the age of LLMs!**

# References

# References i