

# Yerin Hwang

PH.D. CANDIDATE · NLP & LLMs

Bldg. 303, Seoul National University, Seoul, Republic of Korea

✉ dpfls589@snu.ac.kr | in Yerin Hwang

## Summary

Ph.D. candidate at SNU MILAB studying how to evaluate and stress-test Large Language Models. My work examines the robustness of LLM-as-a-Judge and builds more reliable evaluation pipelines. I also explore automated dialogue generation and Korean NLP.

## Education

### Seoul National University

PH.D. CANDIDATE, INTERDISCIPLINARY PROGRAM IN ARTIFICIAL INTELLIGENCE

Seoul, Republic of Korea

Sep. 2021 – Present

### Seoul National University

B.S., DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING (Cum Laude)

Seoul, Republic of Korea

Mar. 2016 – Aug. 2021

## Experience

### LG AI Research

RESEARCH INTERN

Seoul, Republic of Korea

Nov. 2025 – Present

### Max Planck Institute for Security and Privacy (MPI-SP)

RESEARCH INTERN

Bochum, Germany

Aug. 2025 – Oct. 2025

### LG AI Research

INTERN

Seoul, Republic of Korea

Aug. 2022 – Oct. 2022

## Honors & Awards

### National Science & Technology Scholarship

SCHOLARSHIP

Republic of Korea

2018-2019

### SNU IPAI Support Scholarship (Ph.D.)

SCHOLARSHIP

Republic of Korea

### SNU IPAI Support Scholarship (M.S.)

SCHOLARSHIP

Republic of Korea

### Winner of Research Paper Competition, SNU IPAI

AWARDS

Seoul, Republic of Korea

Dec. 2025

### Winner of Research Paper Competition, SNU IPAI

AWARDS

Seoul, Republic of Korea

Jun. 2025

### Winner of Research Paper Competition, SNU IPAI

AWARDS

Seoul, Republic of Korea

Dec. 2024

### Winner of Creative Autonomous Research Competition, SNU IPAI

AWARDS

Seoul, Republic of Korea

Dec. 2023

### Winner of ETRI Human Understanding Artificial Intelligence Paper Contest

AWARDS

Republic of Korea

Jun. 2023

## Publications

---

### **Don't Judge Code by Its Cover: Exploring Biases in LLM Judges for Code Evaluation**

Jiwon Moon\*, **Yerin Hwang\***, Dongryeol Lee, Taegwan Kang, Yongil Kim, Kyomin Jung

- Findings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (**Findings of EACL 2026**)

### **Fooling the LVLM Judges: Visual Biases in LVLM-Based Evaluation**

**Yerin Hwang\***, Dongryeol Lee\*, Kyungmin Min, Taegwan Kang, Yongil Kim, Kyomin Jung

- Conference on Empirical Methods in Natural Language Processing (**EMNLP 2025**)

### **Can You Trick the Grader? Adversarial Persuasion of LLM Judges**

**Yerin Hwang\***, Dongryeol Lee, Taegwan Kang, Yongil Kim, Kyomin Jung

- Findings of Conference on Empirical Methods in Natural Language Processing (**Findings of EMNLP 2025**)

### **LLMs can be easily Confused by Instructional Distractions**

**Yerin Hwang\***, Yongil Kim, Jahyun Koo, Taegwan Kang, Hyunkyung Bae, Kyomin Jung

- The 63rd Annual Meeting of the Association for Computational Linguistics (**ACL 2025**)

### **Are LLM-Judges Robust to Expressions of Uncertainty? Investigating the Effect of Epistemic Markers on LLM-based Evaluation**

Dongryeol Lee\*, **Yerin Hwang\***, Yongil Kim, Joonsuk Park, Kyomin Jung

- Annual Conference of the North American Chapter of the Association for Computational Linguistic (**NAACL 2025, Oral Presentation**)

### **SWITCH: Studying with Teacher for Knowledge Distillation of Large Language Models**

Jahyun Koo, **Yerin Hwang\***, Yongil Kim, Taegwan Kang, Hyunkyung Bae, Kyomin Jung

- Findings of Annual Conference of the North American Chapter of the Association for Computational Linguistic (**Findings of NAACL 2025**)

### **MP2D: An Automated Topic Shift Dialogue Generation Framework Leveraging Knowledge Graphs**

**Yerin Hwang\***, Yongil Kim, Yunah Jang, Jeessoo Bang, Hyunkyung Bae, Kyomin Jung

- Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)

### **Kosmic: Korean Text Similarity Metric Reflecting Honorific Distinctions**

**Yerin Hwang\***, Yongil Kim, Jeessoo Bang, Hyunkyung Bae, Hwanhee Lee, Kyomin Jung

- The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (**COLING 2024**)

### **Dialogizer: Context-aware Conversational-QA Dataset Generation from Textual Sources**

**Yerin Hwang\***, Yongil Kim\*, Hyunkyung Bae, Hwanhee Lee, Jeessoo Bang, Kyomin Jung

- Conference on Empirical Methods in Natural Language Processing (**EMNLP 2023**)

### **PR-MCS: Perturbation Robust Metric for Multilingual Image Captioning**

Yongil Kim, **Yerin Hwang\***, Hyeongu Yun, Seunghyun Yoon, Trung Bui, Kyomin Jung

- Findings of Conference on Empirical Methods in Natural Language Processing (**Findings of EMNLP 2023**)

### **Injecting Comparison Skills in Task-Oriented Dialogue Systems for Database Search Results Disambiguation**

Yongil Kim\*, **Yerin Hwang\***, Joongbo Shin, Hyunkyung Bae, Kyomin Jung

- Findings of Association for Computational Linguistic (**Findings of ACL 2023**)

### **Improving Cross-Modal Attention via Object Detection**

Yongil Kim, **Yerin Hwang\***, Seunghyun Yoon, Hyeongu Yun, Kyomin Jung

- Workshop of The Thirty-Sixth Annual Conference on Neural Information Processing Systems, (**NeurIPS Workshops 2022**)

## INTERNATIONAL JOURNAL

### **Flowlogue: A Novel Framework for Synthetic Dialogue Generation with Structured Flow from Text Passages**

Yongil Kim, Yerin Hwang, Hyunkyung Bae, Taegwan Kang, Kyomin Jung

• IEEE Access (2024)

## DOMESTIC CONFERENCES & JOURNAL

### **A Study on the Evaluation Consistency of Korean LLM-as-a-Judge Models in Mathematical Problems**

Yerin Hwang, Dongryeol Lee, Jiwon Moon, Kyungmin Min, Kyomin Jung

• KCC 2025

### **Analysis of Stylistic Bias in Korean LLM as a Judge**

Yerin Hwang, Dongryeol Lee, Jiwon Moon, Kyungmin Min, Kyomin Jung

• KCC 2025

### **Evaluating the Robustness of LLM-Judges to Epistemic Markers in Korean**

Dongryeol Lee, Yerin Hwang, Jiwon Moon, Kyungmin Min, Kyomin Jung

• KCC 2025

### **TSDG: A Framework for Generating Natural Topic-Shift Dialogue Data**

Yerin Hwang, Dongryeol Lee, Yongil Kim, Kyomin Jung

• KSC 2024

### **Error-Correction Chain of Thought (ECO CoT): Enhancing Accuracy in Mathematical Reasoning through Error-Correction Framework**

Yerin Hwang, Yongil Kim, Dongryeol Lee, Taegwan Kang, Hyunkyung Bae, Kyomin Jung

• KSC 2024

### **Reference-Centric QA Evaluation Leveraging Contrastive Decoding**

Dongryeol Lee, Kyungmin Min, Yerin Hwang, Joonsuk Park, Kyomin Jung

• KSC 2024

### **KLIPScore: A Highly Human-Correlated Korean Image Captioning Metric (Oral)**

Yongil Kim, Yerin Hwang, Yoonhyung Chae, Seunghyun Yoon, Kyomin Jung

• KCC 2023

### **Thinking Fast and Slow in Multimodal Emotion Recognition Task**

Yerin Hwang, Yongil Kim, Yoonhyung Chae, Kyomin Jung

• KCC 2023

### **Improving Cross-Modal Attention via Object Detection**

Yongil Kim, Hyeongu Yun, Yerin Hwang, Kyomin Jung

• KCC 2022

### **COVID-19 Severity Prediction using Deep Transfer Learning**

Yerin Hwang, Yongil Kim, Kyomin Jung

• KCC 2022