Sungmin Kang

Tel: +1 (213) 826-2554 **Email:** kangsung@usc.edu **LinkedIn** **Google Scholar** **Personal Website**

Research Interests

My research interests primarily lie in building **trustworthy and interpretable models**. Specifically:
· **Reliable and Truthful Generation:** I am interested in designing practical methods that quantify LLM response uncertainty and detect factual inconsistencies, remaining robust under distribution shifts and retrieval-augmented settings.
· **Interpretability and Mechanistic Understanding:** I aim to understand the internal mechanisms of large models to analyze and enhance their reasoning, for safe and trustworthy generative process.

Education

University of Southern California (USC) — Los Angeles, United States
M.S. in Electrical Engineering (Advisor: *Prof. Salman Avestimehr*) — Aug. 2024 - May. 2026
GPA: 4.00/4.00 — *MS Honors Fellow*

Sogang University — Seoul, Republic of Korea
B.S. in Electronic Engineering, Micro-degree in Artificial Intelligence — Mar. 2018 - Feb. 2024
GPA: 3.91/4.3 — *Magna Cum Laude*

Publications

1. **GEM: A Scale- and Distribution-Aware Sparse Fine-Tuning Framework for Effective Downstream Adaptation** [PDF]
   Sungmin Kang, Jisoo Kim, Salman Avestimehr, Sunwoo Lee
   The 40th Annual AAAI Conference on Artificial Intelligence (AAAI 2026)
2. **Layer-wise Update Aggregation with Recycling for Communication-Efficient Federated Learning** [PDF]
   Jisoo Kim, Sungmin Kang, Sunwoo Lee
   The Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS 2025)
3. **TruthTorchLM: A Comprehensive Library for Predicting Truthfulness in LLM Outputs** [PDF] [github]
   Duygu Nur Yaldiz*, Yavuz Faruk Bakman*, Sungmin Kang, Alperen Ozis, Hayrettin Eren Yildiz, Mitash Shah, Zhiqi Huang, Anoop Kumar, Alfy Samuel, Daben Liu, Sai Praneeth Karimireddy, Salman Avestimehr
   The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP 2025 System demonstrations)
4. **Reconsidering LLM Uncertainty Estimation Methods in the Wild** [PDF] [Poster] [Slides] [Video]
   Yavuz Faruk Bakman*, Duygu Nur Yaldiz*, Sungmin Kang, Tuo Zhang, Baturalp Buyukates, Salman Avestimehr, Sai Praneeth Karimireddy
   63rd Annual Meeting of the Association for Computational Linguistics (ACL 2025)
5. **Uncertainty as Feature Gaps: Epistemic Uncertainty Quantification of LLMs in Contextual Qusestion-Answering** [PDF]
   Yavuz Faruk Bakman, Sungmin Kang, Zhiqi Huang, Duygu Nur Yaldiz, Catarina G Belém, Chenyang Zhu, Anoop Kumar, Alfy Samuel, Daben Liu, Salman Avestimehr, Sai Praneeth Karimireddy
   Under review at The Fourteenth International Conference on Learning Representations (ICLR 2026)
6. **Uncertainty Quantification for Hallucination Detection in Large Language Models: Foundations, Methodology, and Future Directions**[PDF]
   Sungmin Kang, Yavuz Faruk Bakman, Duygu Nur Yaldiz, Baturalp Buyukates, Salman Avestimehr
   Under review at IEEE BITS the Information Theory Magazine, 2025

Research Experiences

Graduate Research Assistant, University of Southern California — Oct. 2024 - Present
Advisor: *Prof. Salman Avestimehr*, Information Theory and Machine Learning (vITAL) Lab

· Explored the behavior of LLM uncertainty estimation methods across four realistic scenarios: threshold sensitivity, robustness to query transformations, long-form applicability, and ensemble effectiveness. [PDF]
· Proposed a UQ method for contextual QA that measures feature gaps along honesty, context-reliance, and context comprehension via contrastive prompts and representation analysis [PDF]
· Developed an open-source library with 25+ methods for evaluating LLM truthfulness [github] [PDF]

Graudate Research Intern, Inha University — May. 2024 - Present
with *Prof. Sunwoo Lee*, Large-Scale Machine Learning Systems Lab

- Designed a communication-efficient FL algorithm that selectively updates high-variability layers by monitoring gradient-to-weight ratio, reducing communication costs by up to 83% while maintaining model accuracy [PDF]
- Proposed a scale- and distribution-aware fine-tuning method, achieving 1.5% higher accuracy than full fine-tuning with 0.1% tunable parameters via gradient-to-weight ratio and entropy-based masking [PDF]

Undergraduate Research Assistant, Sogang University Jul. 2023 - Apr. 2024
Advisor: *Prof. Hongseok Kim*, Networking for Intelligence Computing and Energy (NICE) Lab

- Built a server-client system upon 8 separate devices through wireless socket communication to implement federated learning, enabling server to handle multiple clients by multi-threading [github]
- Designed a FL algorithm that predicts future gradients via 2D optimization modeling, achieving 1.3x faster convergence and improved accuracy [link]
- Implemented a federated learning algorithm to forecast power in a newly constructed solar power plant in South Korea, deploying the model on four Raspberry Pis

Undergraudate Research Intern, Seoul National University Jan. 2023 - Apr. 2023
with *Prof. Taesup Moon*, Machine Intelligence and Data science (M.IN.D) Lab

- Built a multimodal binary classification model on 1,796 sMRI images by pretraining on age prediction and fine-tuning for MCI-to-AD conversion, achieving an AUROC of 0.84 while addressing label imbalance and data scarcity

Projects

TruthTorchLM: Open-Source Hallucination Detection Library, USC & Capital One Oct. 2024 - May 2025

- Developed an open-source library with 25+ methods for evaluating LLM truthfulness [github] [PDF]
- Provides a unified interface to calibrate, evaluate UQ methods, supporting long-form and contextual QA for both API and HuggingFace models

AI-driven Product Issue Analysis Automation, Sogang University & Mortrex Corporation Aug. 2023 - Dec. 2023

- Automated missing data imputation and symptom classification by developing an AI model for analyzing product issues across global branches of *Motrex Corporation* (*an automotive parts manufacturer*)
- Led a team of four people to predict future trends using the model, and automate effective graph visualizations

OutVenture: In-Car Voice Chatbot, Sogang University & LG Electronics Mar. 2023 - Jun. 2023

- Developed a image classification model to recognize real-time views from moving vehicles in the Unity game engine
- Designed a conversational chatbot that integrates an image classification model to provide real-time information about classified landmarks based on the model outputs [video]

Honors and Awards

- Best Poster Award (out of 110+ teams), ECE 15th Annual Research Festival, USC Oct. 2025
- Best Poster Award (out of 50+ teams), Viterbi PhD Visit Day, USC Apr. 2025
- 3rd Prize (out of 50+ teams), Senior Thesis Project Contest, Sogang University Dec. 2023
- 3rd Prize (out of 40+ teams), Capstone Design Project Contest, Sogang University Jun. 2023
- Daesang Foundation Scholarship (Merit-based scholarship, 5M KRW/semester) Spring 2019 - Fall 2023

Teaching Experiences

- Grader, EE 503: Probability for Engineers (PhD-recommended course) by *Prof. Kosko Bart*, USC Fall 2025
- Teaching Assistant, EEE 4171: AI Communications by *Prof. Hongseok Kim*, Sogang University Spring 2024
- Teaching Assistant, COR 1010: AI Programming by *Prof. Naeun Jang*, Sogang University Summer 2023
- Education Volunteer Services for financially disadvantaged teenage students 2019, 2021

Talks & Conference Presentations

- "Layer-wise Recycling for Communication-Efficient Federated Learning", Sogang University Sep. 2025
- "Reconsidering LLM Uncertainty Estimation Methods in the Wild", Poster Presentation, ACL 2025 [link] Jul. 2025
- "Federated Learning Implementation through Socket Communication", Sogang University [link] May 2024

Leadership Experiences

Sergeant, Korean Augmentation to the US Army (KATUSA), Camp Hovey Oct. 2019 - May 2021

- Led a battalion of 104 KATUSAs as an elected representative, honorably discharged as a Sergeant