Chanwoong Yoon

🌐 cwyoon-99.github.io ✉ cwyoon99@gmail.com 🔗 in/chanwoong-yoon Ⓖ Google Scholar

RESEARCH INTERESTS

Natural Language Processing, AI Safety, Retrieval-Augmented Language Models, Interpretability

EDUCATION

Korea University   Seoul, Republic of Korea
M.S. - Computer Science and Engineering; GPA: 4.0/4.0   [Mar. 2024 - Presents]
Hanyang University   Ansan, Republic of Korea
B.S. - Software, College of Computing; GPA: 3.66/4.0   [Mar. 2017 - Feb. 2023]

ACADEMIC EXPERIENCES

Georgia Institute of Technology   Atlanta, GA, USA
Visiting Researcher (Advisor: Alan Ritter)   [Oct. 2025 - Present]
- Working on fine-grained safety moderation of language models.
- Funded by Korea University–KIAT Scholarship Program (approx. $22,000)
Data Mining and Information Systems Lab, Korea University   Seoul, Republic of Korea
Research Assistant (Advisor: Jaewoo Kang)   [Aug. 2023 - Sep. 2025]
- Improved retriever-augmented language models via context compression and query alignment.
- Developed a benchmark for long-context tasks, controlling for information density.
- Identified specific attention heads in language models for processing time-specific information.
Electronics and Telecommunications Research Institute (ETRI)   Daejeon, Republic of Korea
Research Intern (Advisor: Junseong Bang)   [Jul. 2022 - Aug. 2022]
- Worked on efficient training of Dialogue State Tracking (DST) models.

SELECTED PUBLICATIONS

1. **CompAct: Compressing Retrieved Documents Actively for Question Answering.** C. Yoon, T. Lee, H. Hwang, M. Jeong, J. Kang. EMNLP 2024
2. **Ask Optimal Questions: Aligning Large Language Models with Retriever's Preference in Conversational Search.** C. Yoon*, G. Kim*, B. Jeon, S. Kim, Y. Jo, J. Kang. Findings of NAACL 2025
3. **ETHIC: Evaluating Large Language Models on Long-Context Tasks with High Information Coverage.** T. Lee, C. Yoon, K. Jang, D. Lee, M. Song, H. Kim, J. Kang. NAACL 2025
4. **Temporal Heads: Where Language Models Recall Time-specific Information.** Y. Park, C. Yoon, J. Park, M. Jeong, J. Kang. ACL 2025
5. **ChroKnowledge: Unveiling Chronological Knowledge of Language Models in Multiple Domains.** Y. Park, C. Yoon, J. Park, D. Lee, M. Jeong, J. Kang. ICLR 2025

RESEARCH ACTIVITY

Data Mining and Information Systems Lab, Korea University   Aug. 2023 - Sep. 2025
- CompAct: Context Compression for Retrieved-Augmented Language Model
- Developed an active strategy to compress extensive documents without losing information in context.
- Showed significant improvements in both performance and compression rate on QA benchmarks
- Demonstrated flexibility as a cost-efficient plug-in module with various off-the-shelf retrievers or readers, achieving exceptionally high compression rates (47x).
- RetPO: Retriever's Preference Optimization
- Developed a framework to align query rewriting models with target retriever preferences.

- Achieved superior performance, significantly outperforming existing baselines, including GPT-3.5.
- **Evaluating Language Models on Long-Context Tasks with High Information Coverage**
  - Developed a benchmark to evaluate language models' ability to reason over scattered information in dense, long contexts.

Temporal Heads: Where Language Models Recall Time-Specific Information
  - Discovered Temporal Heads, specific attention heads in language models primarily responsible for processing temporal knowledge.

Electronics and Telecommunications Research Institute (ETRI)   Jul. 2022 - Aug. 2022
- Trained efficient dialogue state tracking models with parameter-efficient fine-tuning

AWARDS AND HONORS

Research Fund Recipient, Korea Institute for Advancement of Technology (KIAT)   May 2025
- Selected to receive over USD 21,000 in research support

Outstanding Research Paper Award, Korea University   Feb. 2025
- CompAct: Compressing Retrieved Documents Actively for Question Answering (EMNLP 2024)

Encouragement Award, KIISE Korea Computer Congress Competition   Jul. 2022

Merit-based Scholarships (50% Tuition), Hanyang University   Fall 2020 — Spring 2022

FULL PUBLICATIONS

6. **Assessing LLM Reasoning Steps via Principal Knowledge Grounding.** H. Hwang, Y. Cho, <u>C. Yoon</u>, Y. Park, M. Song, K. Lee, G. Kim, J. Kang. Findings of EMNLP 2025

7. **Outlier-Safe Pre-Training for Robust 4-Bit Quantization of Large Language Models.** J. Park, T. Lee, <u>C. Yoon</u>, H. Hwang, J. Kang. ACL 2025

8. **Rationale-Guided Retrieval Augmented Generation for Medical Question Answering.** J. Sohn, Y. Park, <u>C. Yoon</u>, S. Park, H. Hwang, M. Sung, H. Kim, J. Kang. NAACL 2025

9. **LAPIS: Language Model-Augmented Police Investigation System.** H. Kim, D. Kim, J. Lee, <u>C. Yoon</u>, D. Choi, M. Gim, J. Kang. CIKM 2024

10. **Small Language Models Learn Enhanced Reasoning Skills from Medical Textbooks.** H. Kim, H. Hwang, J. Lee, S. Park, D. Kim, T. Lee, <u>C. Yoon</u>, J. Sohn, D. Choi, J. Kang. NPJ Digital Medicine

TECHNICAL SKILLS

Programming: Python, C, C++
Libraries: PyTorch, NumPy, Matplotlib, Transformers, TransformerLens
Machine Learning & Deep Learning (LLMs): Post-training Strategies (Fine-tuning, DPO)
Natural Language Processing (NLP): Information Retrieval, Retrieval-Augmented Generation

RESEARCH & ANALYTICAL SKILLS

Theoretical Foundations: Linear Algebra, Probabilistic Modeling, Information Theory
Scientific Writing & Publication: Experience publishing at top-tier conferences (ICLR, ACL, EMNLP, NAACL, CIKM)
Collaboration & Mentorship: Experience in academic collaborations as a first author and co-author (mentoring junior researchers)

ACADEMIC SERVICE

- **Reviewer**: ICLR 2026
- **Secondary Reviewer**: ARR Review (Dec. 2025), ACL 2025

LANGUAGE

Korean (Native)
English (TOEFL: 104)