

# Jaewon Cheon

## RESEARCH INTERESTS

Large Language Models, Efficient Sequence Modeling

## EDUCATION

<b>Korea University</b> <i>M.S., Industrial and Management Engineering (Advisor: Prof. <a href="#">Pilsung Kang</a>)</i> · Thesis: Leveraging Intrinsic Sparsity of GLU-based Large Language Models for Inference Efficiency	Mar 2024 - Feb 2026 GPA: 4.5/4.5
<b>Korea University</b> <i>Bachelor of Media and Communication</i>	Mar 2016 - Feb 2023 GPA: 4.38/4.5

## AWARDS & GRANTS

<b>Basic Science Research Program</b>   <i>Granted by National Research Foundation of Korea</i> · Inference Acceleration of Large Language Models via Sparsity Prediction	Sep 2024 - Aug 2025
--	---------------------

## PUBLICATIONS

[1] <b>COUNTDOWN: Contextually Sparse Activation Filtering Out Unnecessary Weights in Down Projection</b> Jaewon Cheon and Pilsung Kang. <i>Empirical Methods in Natural Language Processing(EMNLP) Main Track, 2025</i>
--

## CONFERENCE PRESENTATIONS

<b>Enhancing Manufacturing Reliability Through NLP-Based Log Anomaly Detection</b> · Presented at the <i>2025 INFORMS Conference on Quality, Statistics and Reliability (ICQSR), Singapore</i>
---

## WORK EXPERIENCE

<b>Tesser Inc.</b>   <i>Natural Language Processing</i> · Research Intern · Topic: Keyword extraction, Medical LLM · Contributed to <a href="#">Ontol</a> @ Tesser	Oct 2023 - Jan 2024
---	---------------------

## PROJECTS

<b>Train Fast &amp; Infer Fast</b>   <i>Ministry of Science and ICT</i> · Implemented a self-distillation approach for transforming full-attention LLMs into hybrid RNN-attention models	Sep 2025 - Feb 2026
<b>Generative AI-based Workflow Automation</b>   <i>Samsung Fire &amp; Marine Insurance</i> · Built hierarchical data conversion pipeline for insurance policy documents for RAG-based workflow automation.	Apr 2025 - Mar 2026
<b>Researcher Matching System</b>   <i>Seoul National University</i> · Developed a RAG-based recommendation system that matches industry partners with suitable researchers	Dec 2024 - Jun 2025
<b>Development of a News-Domain Specialized RAG Pipeline</b>   <i>BECUAI</i> · Enhancing search quality and answer reliability through query refinement and context-aware response structuring	Apr 2024 - Jun 2024