

Cheonbok Park

Research Scientist, NAVER Cloud

cbok.park@navercorp.com | [cbokpark.github.io](https://github.com/cbokpark) | +82 (0)10 2013 0723

chunbok94@gmail.com | cbok.park@kaist.ac.kr

Updated: 24 October 2026

Research Interests

Multilingual large language models, reasoning and reliability, machine translation, and human-in-the-loop machine learning.

Education

Ph.D. in Artificial Intelligence Sep. 2024–Present

Korea Advanced Institute of Science and Technology (KAIST)

Advisor: Prof. Jaegul Choo.

Pursuing the Ph.D. alongside a full-time role at NAVER Cloud, with support from the company.

M.S. in Computer Science 2018–2020

Korea University

Advisor: Prof. Jaegul Choo.

B.S. in Computer Science and Engineering 2013–2018

Korea University

Second major: Statistics.

Work Experience

NAVER Cloud Apr. 2020–Present

Research Scientist

Foundation Research, Hyperscale AI

Research on multilingual large language models, reasoning, and reliability, with contributions to HyperCLOVA X.

Papago 2020–Nov. 2023

Machine translation research; alternative military service.

AI Language Lab, NCSoft Sep. 2017–Feb. 2018

Research Intern | Seongnam, Korea

Technical Director team member.

KLUE Feb. 2017–Sep. 2017

Engineer & Project Director | Seoul, Korea

Honors and Awards

Gold Prize (1st Place), KU Graduate Project Competition 2017

Korea University

Project: Visualizing and interpreting deep learning models.

Republic of Korea Presidential Science Scholarship 2013–2017

Government of the Republic of Korea

Full-tuition scholarship with stipend for undergraduate studies in computer science. One of 135 awardees in Korea.

Publications

* Equal contribution | † Corresponding author.

Conference, Journal, and Workshop Papers

2026

1. **A Simple Unigram Cross-Entropy Lens on the Lexical Imprint of Pre-training Data**
Jeonghoon Kim*, Woojin Chung*, Woomin Song, **Cheonbok Park**, Nako Sung, and Jinwoo Shin
Conference on Neural Information Processing Systems (NeurIPS), 2026. Accepted (Poster). [\[Paper\]](#)
2. **From Unverifiable Domain Text to Cross-Lingual RLVR via Code-Switched Task Synthesis**
Cheonbok Park*, Jihwan Kim*, Shinyoung Joo, Joosung Lee, Hwiyeol Jo, Jeonghoon Kim, and Jaegul Choo
Empirical Methods in Natural Language Processing (EMNLP), 2026. Accepted. [\[Paper\]](#)
3. **What Models Know, How Well They Know It: Knowledge-Weighted Fine-Tuning for Learning When to Say "I Don't Know"**
Joosung Lee, Hwiyeol Jo, Donghyeon Ko, Kyubyeung Chae, **Cheonbok Park**, and Jeonghoon Kim
Findings of the Association for Computational Linguistics: EMNLP, 2026. Accepted. [\[Paper\]](#)
4. **Enhancing Hallucination Detection via Future Context**
Joosung Lee, **Cheonbok Park**, Hwiyeol Jo, Jeonghoon Kim, Joonsuk Park, and Kang Min Yoo
Findings of the Association for Computational Linguistics: ACL, 2026. [\[Paper\]](#)
5. **Sommelier: Scalable Open Multi-turn Audio Pre-processing for Full-duplex Speech Language Models**
Kyudan Jung, Jihwan Kim, Soyeon Kim, Jeonghoon Kim, Jaegul Choo, and **Cheonbok Park**†
Annual Meeting of the Association for Computational Linguistics (ACL), Industry Track, 2026. [\[Paper\]](#)

2025

6. **Code-Switching Curriculum Learning for Multilingual Transfer in LLMs**
Haneul Yoo, **Cheonbok Park**, Sangdoo Yun, Alice Oh, and Hwaran Lee
Findings of the Association for Computational Linguistics: ACL, 2025. [\[Paper\]](#)
7. **Peri-LN: Revisiting Normalization Layer in the Transformer Architecture**
Jeonghoon Kim, Byeongchan Lee, **Cheonbok Park**, Yeontaek Oh, Beomjun Kim, Taehwan Yoo, Seongjin Shin, Dongyoon Han, Jinwoo Shin, and Kang Min Yoo
International Conference on Machine Learning (ICML), 2025. [\[Paper\]](#)
8. **KMMLU: Measuring Massive Multitask Language Understanding in Korean**
Guijin Son, Hanwool Lee, Sungdong Kim, Seungone Kim, Niklas Muennighoff, Taekyoon Choi, **Cheonbok Park**, Kang Min Yoo, and Stella Biderman
Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL), Long Paper, 2025. [\[Paper\]](#)

2024

9. **Aligning Language Models to Explicitly Handle Ambiguity**
Hyuhng Joon Kim, Youna Kim, **Cheonbok Park**, Junyeob Kim, Choonghyun Park, Kang Min Yoo, Sang-goo Lee, and Taeuk Kim
Empirical Methods in Natural Language Processing (EMNLP), 2024. [\[Paper\]](#)
10. **Adaptive Contrastive Decoding in Retrieval-Augmented Generation for Handling Noisy Contexts**
Youna Kim, Hyuhng Joon Kim, **Cheonbok Park**, Choonghyun Park, Hyunsoo Cho, Junyeob Kim, Kang Min Yoo, Sang-goo Lee, and Taeuk Kim
Findings of the Association for Computational Linguistics: EMNLP, 2024. [\[Paper\]](#)
11. **Mitigating Semantic Leakage in Cross-lingual Embeddings via Orthogonality Constraint**
Dayeon Ki, **Cheonbok Park**, and Hyunjoong Kim
Workshop on Representation Learning for NLP (RepL4NLP), 2024. [\[Paper\]](#)

2023

12. **Towards Accurate Translation via Semantically Appropriate Application of Lexical Constraints**
Yujin Baek*, Koanho Lee*, Dayeon Ki, **Cheonbok Park**, Hyoung-Gyu Lee, and Jaegul Choo
Annual Meeting of the Association for Computational Linguistics (ACL), Findings of ACL, 2023. [Paper]
13. **PePe: Personalized Post-editing Model utilizing User-generated Post-edits**
Jihyeon Lee*, TaeHee Kim*, Yunwon Tae*, **Cheonbok Park**, and Jaegul Choo
Conference of the European Chapter of the Association for Computational Linguistics (EACL), Findings of EACL, 2023. [Paper]
14. **A Visual Analytics System for Improving Attention-based Traffic Forecasting Models**
Seungmin Jin, Hyunwook Lee, **Cheonbok Park**, Hyesin Chu, Yunwon Tae, Jaegul Choo, and Sungahn Ko
IEEE Trans. on Visualization and Computer Graphics (TVCG), 2023 (Proc. IEEE VIS'22) [Paper]

2022

15. **Specializing Multi-domain NMT via Penalizing Low Mutual Information**
Jiyoung Lee, Hantae Kim, Hyunchang Cho, Edward Choi, and **Cheonbok Park**[†]
Empirical Methods in Natural Language Processing (EMNLP), Short Paper, 2022 [Paper]
16. **Residual Correction in Real-Time Traffic Forecasting**
Daejin Kim*, Youngin Cho*, Dongmin Kim, **Cheonbok Park**, and Jaegul Choo
ACM International Conference on Information and Knowledge Management (CIKM), 2022. [Paper]
17. **DaLC: Domain Adaptation Learning Curve Prediction for Neural Machine Translation**
Cheonbok Park, Hantae Kim, Ioan Calapodescu, Hyun Chang Cho, and Vassilina Nikoulina
Meeting of the Association for Computational Linguistics (ACL), Findings Long Paper, 2022 [Paper]
18. **Reversible Instance Normalization for Accurate Time-Series Forecasting against Distribution Shift**
Taesung Kim*, Jinhee Kim*, Yunwon Tae, **Cheonbok Park**, Jang-Ho Choi, and Jaegul Choo
International Conference on Learning Representations (ICLR), 2022. [Paper]
19. **PASTA: PARallel Spatio-Temporal Attention with Spatial Auto-Correlation Gating for Fine-Grained Crowd Flow Prediction**
Chung Park, Junui Hong, **Cheonbok Park**, Taesan Kim, Minsung Choi, and Jaegul Choo
Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2022. [Paper]

2021

20. **Unsupervised Neural Machine Translation for Low-Resource Domains via Meta-Learning**
Cheonbok Park*, Yunwon Tae*, TaeHee Kim, Soyoung Yang, Mohammad Azam Khan, Lucy Park and Jaegul Choo
Meeting of the Association for Computational Linguistics (ACL), Long Paper, 2021 [Paper]
21. **VATUN: Visual Analytics for Testing and Understanding Convolutional Neural Networks**
Cheonbok Park*, Soyoung Yang*, Inyoun Na*, Sunghyo Chung, Sungbok Shin, Bum Chul Kwon, Deokgun Park, and Jaegul Choo
EG/VGTC Conference on Visualization (EuroVis), Short Paper, 2021 [Paper]
22. **An Empirical Experiment on Deep Learning Models for Predicting Traffic Data**
Hyunwook Lee, **Cheonbok Park**, Seungmin Jin, Hyesin Chu, Jaegul Choo, and Sungahn Ko
IEEE International Conference on Data Engineering (ICDE), Short Paper, 2021. [Paper]

2020

23. **ST-GRAT: A Novel Spatio-temporal Graph Attention Networks for Accurately Forecasting Dynamically Changing Road Speed**
Cheonbok Park, Chunggi Lee, Hyojin Bahng, Yunwon Tae, Seungmin Jin, Kihwan Kim, Sungahn Ko, and Jaegul Choo
ACM International Conference on Information and Knowledge Management (CIKM), 2020 [Paper]

2019

24. SANVis: Visual Analytics for Understanding Self-Attention Networks

Cheonbok Park, Inyoun Na, Yongjang Jo, Sungbok Shin, Jaehyo Yoo, Bum Chul Kwon, Jian Zhao, Hyungjong Noh, Yeonsoo Lee, and Jaegul Choo

IEEE VIS, Short Paper, 2019, Vancouver, Canada. [Paper]

25. AILA: Attentive Interactive Labeling Assistant for Document Classification through Attention-based Deep Neural Networks

Minsuk Choi, Cheonbok Park, Soyoung Yang, Yonggyu Kim, Jaegul Choo, and Sungsoo (Ray) Hong

ACM CHI Conference on Human Factors in Computing Systems (CHI), 2019, Glasgow, UK. [Paper]

26. A Comparison of the Effects of Data Imputation Methods on Model Performance

Wooyoung Kim*, Wonwoong Cho*, Jangho Choi, Jiyong Kim, Cheonbok Park, Jaegul Choo

IEEE International Conference on Advanced Communications Technology (ICACT), 2019, Pyeongchang, KR. [Paper]

2017

27. Real-time UAV sound detection and analysis system

Juhyun Kim, Cheonbok Park, Jinwoo Ahn, Youlim Ko, Junghyun Park, John C. Gallagher

IEEE Sensors Applications Symposium (SAS), 2017, Glassboro, NJ, USA. [Paper]

2016

28. ReVACNN: Steering Convolutional Neural Network via Real-Time Visual Analytics

Sunghyo Chung, Cheonbok Park, Sangho Suh, Kyeongpil Kang, Jaegul Choo, and Bum Chul Kwon

NIPS Workshop on the Future of Interactive Machine Learning (FILM), 2016. [Program]

29. ReVACNN: Real-Time Visual Analytics for Convolutional Neural Network

Sunghyo Chung, Sangho Suh, Cheonbok Park, Kyeongpil Kang, Jaegul Choo, and Bum Chul Kwon

ACM SIGKDD Workshop on Interactive Data Exploration and Analytics (KDD-IDEA), 2016, San Francisco, CA. [Paper]

Preprints (Under Review)

30. When Verifiable Rewards Switch the Language: Cross-Lingual Collapse in Chain-of-Thought

Cheonbok Park*, Jeonghoon Kim*, Joosung Lee, Sanghwan Bae, Jaegul Choo, and Kang Min Yoo

arXiv preprint, 2025 (revised February 2026). Under review. [Paper]

31. ReGUIDE: Data Efficient GUI Grounding via Spatial Reasoning and Search

Hyunseok Lee, Jeonghoon Kim, Beomjun Kim, Jihoon Tack, Chansong Jo, Jaehong Lee, Cheonbok Park, Sookyo In, Jinwoo Shin, and Kang Min Yoo

arXiv preprint, 2025. Under review. [Paper]

Technical Reports

32. HyperCLOVA X 32B Think

NAVER Cloud HyperCLOVA X Team (including Cheonbok Park)

Technical report, 2026. [Paper]

33. HyperCLOVA X 8B Omni

NAVER Cloud HyperCLOVA X Team (including Cheonbok Park)

Technical report, 2026. [Paper]

34. HyperCLOVA X THINK Technical Report

NAVER Cloud HyperCLOVA X Team (including Cheonbok Park)

Technical report, 2025. [Paper]

35. HyperCLOVA X Technical Report

Kang Min Yoo et al. (including Cheonbok Park)

Technical report, 2024. [Paper]

Research Experience

Data and Visual Analytics Lab

Mar. 2018–Feb. 2019

Research Intern | Seoul, Korea

Advisor: Jaegul Choo. Research in deep learning, natural language processing, and generative models.

IITP Purdue Junior Capstone Program, Purdue University

Sep. 2016–Feb. 2017

Machine Learning Engineer & Research Intern | West Lafayette, IN

Advisors: John C. Gallagher and Eric T. Matson.

Data and Visual Analytics Lab

Jan. 2016–Sep. 2017

Research Intern | Seoul, Korea

Advisor: Jaegul Choo. Research in deep learning, interpretable machine learning, and visualization.

Talks

Meta-Learning for the Low-Resource Domain

2021

NAVER Search Colloquium.

데이터, 정말 묻고 더블로 가? : 기계번역 성능 개선 예측 모델

2021

DEVIEW.

Teaching

Deep Learning (SAMSUNG DS), KAIST

Fall 2019

Teaching: convolutional neural networks and attention models.

Big Data Analysis and Practice (DFC609), Korea University

Spring & Fall 2019

Teaching Assistant.

Data Science: Deep Learning, Korea University

Aug. 2019

Teaching: training neural networks, recurrent neural networks, and attention models.

Deep Learning (COSE474), Korea University

Fall 2018

Teaching Assistant.

Data Science: Deep Learning, Korea University

Aug. 2018

Teaching: generative models, training neural networks, and convolutional neural networks.

Deep Learning, SKT

May 2018

Lab Teaching Assistant.

Fundamental Mathematics for Machine Learning, Korea University

Feb. & May 2018

Calculus, probability, and linear algebra.

Technical Skills and Languages

Programming and tools: Python, PyTorch, SQL, R, D3.js, HTML, C, PHP, and $\text{\LaTeX}$.

Languages: Korean (native); English (intermediate).