**Boseop Kim**

Seoul, South Korea • (+82) 10 6405 6417

[bsk0130@gmail.com](mailto:bsk0130@gmail.com) • [github.com/seopbo](https://github.com/seopbo) • velog.io/@aisolab

**RESEARCH INTEREST**

* Natural Language Processing, Natural Language Understanding, Information Retrieval, Conversation
* Pre-trained language model, Extreme large scale pre-trained language model

**EDUCATION**

**Korea University** Seoul, South Korea, Sep 2015 – Aug 2017

* M.S in Industrial Management Engineering
* GPA 4.5, Full Marks: 4.5
* Honors student for several semesters
* Relevant Coursework: Machine Learning, Unstructured Data Analysis, Applied Statistics Methods

**Konkuk University** Seoul, South Korea, Mar 2008 – Aug 2015

* B.S. in Applied Statistics
* GPA 4.11, Full Marks: 4.5
* Honors student for several semesters
* Relevant Coursework: Data Mining, Multivariate Data Analysis, Bayesian Statistics

**INDUSTRY EXPERIENCE**

**NAVER** Gyeonggi-do, South Korea, Mar 2020 – Current

* Research Engineer
* Implemented various scale of KoGPT3 models in HyperCLOVA project
* Implemented a data preprocessing pipeline in HyperCLOVA project
* Implemented various pre-trained language models (e.g RoBERTa, GPT2)
* Implemented various chatbot models of NAVER Chatbot service

**LG Electronics** Seoul, South Korea, Aug 2018 – Feb 2020

* Research Engineer
* Implemented modules for quantifying uncertainties of the intent, domain classification model in Spoken Language Understanding System
* Implemented the intent classification model using the pretrained BERT model

**KT** Seoul, South Korea, Jun 2017 – Jul 2018

* Research Engineer
* Implemented the model of predicting transmission network failure using the convolution neural network

**RESEARCH EXPERIENCE**

**Data Science & Business Analytics Lab** Seoul, South Korea, Sep 2015 – Aug 2017

* Researcher (Adviser: Pilsung Kang)
* Researched into Text Mining, Natural Language Processing, Clustering
* Thesis: Integrating cluster validity indices based on data envelopment analysis

**PUBLICATIONS**

* ***Kim, Boseop*** et al. “What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA: Billions-scale Korean Generative Pretrained Transformers.” EMNLP (2021).
* ***Kim, Boseop*** et al. “Integrating cluster validity indices based on data envelopment analysis.” Appl. Soft Comput. 64 (2018): 94-108.

**PROJECTS**

**LASSL** Jan 2021 – Current

* Implemented a framework for training language models (under construction)  
  (project page: <https://github.com/lassl/lassl>)

**NLP Implementation**  Jan 2019 – Jul 2019

* Implemented benchmark papers of Natural Language Processing using Korean Corpus  
  (project page: <https://github.com/seopbo/nlp_implementation>)

**Deep Learning for All Season 2** Aug 2018 – Mar 2019

* Contributed to lectures of recurrent neural network using TensorFlow

(project page: <https://deeplearningzerotoall.github.io/season2/lec_tensorflow.html>)

**Fault Detection of semiconductor wafers** Feb 2018 – Jun 2019

* Implemented the classification model using the convolution neural network for benchmark
* Implemented the module of interpretation for the convolution neural network

**EXTRACURRICULAR ACTIVITIES**

**NLP Bootcamp**  Oct 2018 – Feb 2019

* Contributed the study group for Natural Language Understanding, Natural Language Processing as a facilitator (project page: <https://github.com/modulabs/NLP-bootcamp>)

**DeepNLP** Jan 2019 – Jan 2020

* Reviewed a lot of papers of Natural Language Processing
* Contributed the mentor program as a mentor

**PRESENTATIONS**

**NAVER DEVIEW**  NAVER DEVIEW, Nov 2021

* Gave a presentation for Multilingual HyperCLOVA

**NAVER AI NOW**  NAVER AI NOW, May 2021

* Gave a presentation for HyperCLOVA’s KoGPT3

**Starting implementation of Natural Language Processing papers**  PyCon KR, Oct 2019

* Gave a tutorial for implementing papers of Natural Language Processing

**Structuring your first NLP Project** Deep Learning Conference All Together**,** Oct 2019

* Gave a presentation for basic project structure of Deep Learning project

**HONORS & AWARD**

* Excellence award, 5th Industrial fusion revitalization plan and case studies contest, KIIE, 2016
* Participation award, 1st Bigdata Analytics Festival, SKT, 2015

**TECHNICAL SKILLS**

**Languages**

* Python ●●●●○
* Scala ●●○○○

**Frameworks / Libraries**

* PyTorch ●●●●○
* TensorFlow ●●●○○
* PySpark ●●●○○
* Spark ●●○○○

**Environments / ETC**

* Git ●●●○○
* Docker ●●●○○
* Linux ●●●○○
* Vscode ●●●○○