Joel Jang

Last updated on May 22, 2023

joeljang.github.io ♦ joeljang@kaist.ac.kr

RESEARCH INTERESTS

My main research goal is to build Large Language Models (LLMs) that are applicable to real-world scenarios, by addressing inherent limitations of current LLMs. Specifically, I am interested in allowing LLMs to be lifelong learners [C1, C2, C5], providing privacy and security guarantees [C7], and enabling LLMs to follow the given instruction via prompt [W1, C3, C4, P1].

EDUCATION

Ph.D. in Computer Science *Seattle, US*
University of Washington 09/2023 -
Advisors: Luke Zettlemoyer

M.S. in Artificial Intelligence *Seoul, Korea*
Korea Advanced Institute of Science and Technology (KAIST) 03/2021 - *08/2023*
Advisor: Minjoon Seo

B.S. in Computer Science *Seoul, Korea*
Korea University 03/2017 - 02/2021

PUBLICATIONS

Conference Papers

[C7] Knowledge Unlearning for Mitigating Privacy Risks in Language Models
Joel Jang, Dongkeun Yoon, Sohee Yang, Sungmin Cha, Moontae Lee, Lajanugen Logeswaran, Minjoon Seo
ACL 2023 [paper][code]

[C6] Gradient Ascent Post-training Enhances Language Model Generalization
Dongkeun Yoon*, **Joel Jang***, Sungdong Kim, Minjoon Seo
ACL 2023

[C5] Prompt Injection: Parameterization of Fixed Inputs
Eunbi Choi, Yongrae Jo, **Joel Jang**, Joonwon Jang, Minjoon Seo
ACL 2023 Findings [paper][code]

[C4] Exploring the Benefits of Training Expert Language Models over Instruction Tuning
Joel Jang, Seungone Kim, Seonghyeon Ye, Doyoung Kim, Lajanugen Logeswaran, Moontae Lee, Kyungjae Lee, Minjoon Seo
ICML 2023 [paper][code]

[C3] Guess the Instruction! Making Language Models Stronger Zero-shot Learners
Seonghyeon Ye, Doyoung Kim, **Joel Jang**, Joongbo Shin, Minjoon Seo
ICLR 2023 [paper][code]

[C2] TemporalWiki: A Lifelong Benchmark for Training and Evaluating Ever-Evolving Language Models
Joel Jang*, Seonghyeon Ye*, Changho Lee, Sohee Yang, Joongbo Shin, Janghoon Han, Gyeonghun Kim, Minjoon Seo
EMNLP 2022 [paper][code]

[C1] Towards Continual Knowledge Learning of Language Models
Joel Jang, Seonghyeon Ye, Sohee Yang, Joongbo Shin, Janghoon Han, Gyeonghun Kim, Stanley Jungkyu Choi, Minjoon Seo
ICLR 2022 [paper] [code]

Workshop Papers

[W1] Can Large Language Models Truly Follow Your Instructions? Case-study with Negated Prompts
Joel Jang*, Seonghyeon Ye*, Minjoon Seo
NeurIPS 2022 Workshop on Transfer Learning for NLP (TL4NLP) [paper][code]

Journal Papers

[J2] Sequential Targeting: A Continual Learning Approach for Data Imbalance in Text Classification
Joel Jang, Yoonjeon Kim, Kyoungho Choi, Sungho Suh
Expert Systems With Applications (2021) [paper] [code]

[J1] Supervised Health Stage Prediction Using Convolution Neural Networks for Bearing Wear
Sungho Suh, **Joel Jang**, Seungjae Won, Mayank S. Jha, Yong Oh Lee
Sensors (2020) [paper] [code]

Preprints

[P1] Retrieval of Soft Prompt Enhances Zero-shot Task Generalization
Seonghyeon Ye, **Joel Jang**, Doyoung Kim, Yongrae Jo, Minjoon Seo
Under Review [paper][code]

EXPERIENCE

Allen Institute of AI (AI2) | Mosaic Team *Seattle, US*
Research Intern (Host: Prithviraj (Raj) Ammanabrolu) *06/2023-*
Personalized Reinforcement Learning from Human Feedback (RLHF)

LG AI Research *Seoul, Korea*
Research Intern (Host : Moontae Lee, Lajanugen Logeswaran*)* 07/2022-05/2023
Working on (1) knowledge unlearning for LMs & (2) unseen task generalization with expert LMs

Kakao Brain *Seongnam, Korea*
Research Intern (Host : Ildoo Kim *)* 12/2020-02/2021
Worked on large-scale representation learning with weak supervision of images and caption data using TPUs

NAVER *Seongnam, Korea*
Software Engineer Intern 07/2020-09/2020
Worked on hate speech detection model, AI Clean Bot 2.0 (40+ million monthly users, >80% of Korean population)
Developed novel method of handling data imbalance using continual learning (*paper published under ESWA*)

KIST Europe *Saarbrucken, Germany*
Research Intern (Host: Sungho Suh *)* 08/2019-01/2020
Worked on anomaly detection & remaining useful life prediction of machinery (*paper published under Sensors*)
Gave an Oral Presentation at *PHM Korea 2020* (2020. 07. 23)

SERVICES

Conference Reviewer
COLING 2022, EMNLP 2022, AKBC 2022, ICLR 2023, ACL 2023

Journal Reviewer
Journal of Artificial Intelligence Research (JAIR)

MENTORING

06/2022-05/2023 **Dongkeun Yoon**, B.S. Student, Konkuk University → M.S./Ph.D. Student, KAIST
06/2022-05/2023 **Seungone Kim**, B.S. Student, Yonsei → M.S. Student, KAIST
10/2021-05/2022 **Changho Lee**, B.S. Student, Korea University → LG AI Research
06/2021-02/2022 **Seonghyeon Ye**, B.S. Student, KAIST → M.S./Ph.D. Student, KAIST

TEACHING

(KAIST AI620) NLP Bias and Ethics 03/2023-06/2023
Teaching Assistant (TA) | Instructor: James Throne

(KAIST AI599) AI for Law 09/2022-12/2022
Teaching Assistant (TA) | Instructor: Minjoon Seo

(KAIST AI605) Deep Learning for NLP 03/2022-06/2022
Teaching Assistant (TA) | Instructor: Minjoon Seo

INVITED TALKS

Continual Learning for Language Models 04/2023
ContinualAI (Host: James Smith)

Expert Language Models 02/2023
UNC at Chapel Hill (Host: Colin Raffel)

Temporal Adaptation of Language Models 08/2022
Korean AI Association Summer NLP Session (Host: Minjoon Seo)

Temporal Adaptation of Language Models 07/2022
KAIST School of Computing (Host: Alice Oh)

Temporal Adaptation of Language Models 05/2022
Hyperconnect (Host: Buru Chang)

HONORS AND AWARDS

Qualcomm Innovation Fellowship Korea (QIFK) 2022
Grand Prize in Graduation Capstone Competition (Best Paper Award), 2020 (*Advisor:* Jaewoo Kang*)*
4th place, AI NLP Challenge Enliple Cup, 2020
3rd place, HAAFOR Challenge 2019

Future Global Leader Scholarships, Korea University, 2019
Best Innovation Award, Intel AI Drone Hackathon, 2018

LANGUAGE PROFICIENCY

Bilingual in English (*2004-2016 in US*) and Korean (*native)*
GRE: 326 (Verbal, 157/170, 76th Percentile) | Quant, 169/170, 95th Percentile | Writing, 5.0/6.0, 92nd Percentile)
TOEFL: 119/120 (Reading, 30 | Listening, 30 | Speaking, 29 | Writing, 30)
SAT: 1530/1600 (Reading and Writing, 730 | Math, 800)
Conversational in Chinese