Joel Jang

Last updated on Feb 6, 2024

joeljang.github.io ♦ joeljang@cs.washington.edu

EDUCATION

Ph.D. in Computer Science — Seattle, US
University of Washington — 09/2023 -
Advisor: Luke Zettlemoyer, Dieter Fox

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 and Engineering — Seoul, Korea
Korea University — 03/2017 - 02/2021

PUBLICATIONS

Preprints

[P5] Personalized Soups: Personalized LArge Language Model Alignment via Post-hoc Parameter Merging
Joel Jang, Seungone Kim, Bill Yuchen Lin, Yizhong Wang, Jack Hessel, Luke Zettlemoyer, Hannaneh Hajishirzi, Yejin Choi, Prithviraj Ammanabrolu
Under Review

[P4] Camels in a Changing Climate: Enhancing LM Adaptation with Tulu 2
Hamish Ivison, Yizhong Wang, Valentina Pyatkin, Nathan Lambert, Matthew Peters, Pradeep Dasigi, **Joel Jang**, David Wadden, Noah A. Smith, Iz Beltagy, Hannaneh Hajishirzi
Under Review

[P3] LangBridge: Multilingual Reasoning Without Multilingual Supervision
Dongkeun Yoon, **Joel Jang**, Seongone Kim, Sheikh Shafayat, Minjoon Seo
Under Review

[P2] How Well Do Large Language Models Truly Ground?
Hyunji Lee, Sejune Joo, Chaeeun Kim, **Joel Jang**, Doyoung Kim, Kyoungwoon On, Minjoon Seo
Under Review

[P1] Continually Updating Generative Retrieval on Dynamic Corpora
Soyoung Yoon, Chaeeun Kim, Hyunji Lee, **Joel Jang**, Minjoon Seo
Under Review

Conference Papers

[C10] Prometheus: Inducing Fine-grained Evaluation Capability in Language Models
Seungone Kim*, Jamin Shin*, Yejin Cho*, **Joel Jang**, Shayne Longpre, Hwaran Lee, Sangdoo Yun, Seongjin Shin, Sungdong Kim, James Thorne, Minjoon Seo
ICLR 2024

[C9] The CoT Collection: Improving Zero-shot and Few-shot Learning of Language Models via Chain-of-Thought Fine-Tuning
Seungone Kim*, Se June Joo*, Doyoung Kim, **Joel Jang**, Seonghyeon Ye, Jamin Shin, Minjoon Seo
EMNLP 2023

[C8] Retrieval of Soft Prompt Enhances Zero-shot Task Generalization
Seonghyeon Ye, **Joel Jang**, Doyoung Kim, Yongrae Jo, Minjoon Seo

EMNLP 2023 Findings [paper] [code]

[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

[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

[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

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

[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

[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

Conference 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)

Journal Papers

[J3] Improving Probability-based Prompt Selection Through Unified Evaluation and Analysis
Sohee Yang, Jonghyeon Kim, **Joel Jang**, Seonghyeon Ye, Hyunji Lee, Minjoon Seo
TACL 2024

[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)

[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)

EXPERIENCE

Allen Institute of AI (Ai2) Seattle, US
Research Intern (Mentors: : Prithviraj Ammanabrolu, Yejin Choi *)* 06/2023 - 01/2024
(1) Personalized RLHF of LLMs and (2) Democratizing Robotic Foundational Models

LG AI Research Seoul, Korea

Research Intern (Mentors : Moontae Lee, Lajanugen Logeswaran *)* 07/2022 - 05/2023

(1) Knowledge Unlearning for LLMs and (2) Expert Language Models

Kakao Brain Seongnam, Korea

Research Intern (Mentor : Ildoo Kim *)* 12/2020 - 02/2021

Representation learning of vision-language models with weakly supervision

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 (Mentor : 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)

MENTORING

KAIST AI

Dongkeun Yoon (2022 - 2023), Konkuk University Undergrad. Published [C6, C7] Now Ph.D. at KAIST AI
Seungone Kim (2022 - 2023), Yonsei University Undergrad. Published [C4]. Now Ph.D. at CMU
Changho Lee (2021 - 2022), Korea University Undergrad. Published [C2]. Now Research Scientist at LG AI Research
Seonghyeon Ye (2021 - 2022), KAIST Undergrad. Published [C1,C2,W1]. Now Ph.D. at KAIST AI

SERVICES

Conference Reviewer

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

Journal Sub-reviewer

Journal of Artificial Intelligence Research (JAIR)

TEACHING

(KAIST AI620) NLP Bias and Ethics *Spring 2023*

Teaching Assistant (TA) | Instructor: James Throne

(KAIST AI599) AI for Law *Fall 2022*

Teaching Assistant (TA) | Instructor: Minjoon Seo

(KAIST AI605) Deep Learning for NLP *Spring 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, $2000), 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