

Jia-Huei (Dylan) Ju

Research Center of Information Technology Innovation, Academia Sinica
128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan (R.O.C)
<https://dylanjoo.github.io> | <https://github.com/DylanJoo>

+886-983766845
jhjoo@citi.sinica.edu.tw

EDUCATION

National Chengchi University

M.Sc. in Management Information Systems

- Overall GPA: 4.17/4.3
- Thesis: High-Dimensional VAR for Retail Marketing and Sales Performance Analysis (advisors: Prof. Hao-Chun Chuang and Prof. Yen-Chun Chou)

Taipei, Taiwan
Sep. 2019 - Sep. 2021

National Central University

B.B.A in Information Management

- Overall GPA: 3.09/4.0

Taoyuan, Taiwan
Sep. 2015 - June 2019

RESEARCH EXPERIENCE

The CFDA Lab, Academia Sinica

Part-time/ Full-time Research Assistant | Supervisor: Prof. Chuan-Ju Wang

Taipei, Taiwan
Oct. 2019 - Dec. 2021/
May 2022 - Present

Neural Information Retrieval: Improving Passage Re-ranking with Text Generation

- Designed text-to-text multi-view learning framework for passage re-ranking; the multi-view model using 770M parameters outperformed single-view model using 3B parameters.
- Published a short paper and presented the work at SIGIR'21.

Neural Information Retrieval: Domain-adaptive Few-shot Dense Retrieval

- Constructed relevance-aware question generators with instruction-tuning and knowledge distillation.
- Retrofitted generalized dense retrievers into domain-adaptive retriever via few-shot learning.

Conversational Search: Weakly-supervised Conversational Passage Re-ranking

- Constructed a view ensemble pseudo-labeling approach for training conversational passage re-ranking model; the proposed model improved supervised learning baseline method by 10% nDCG@3 with lower query latency.
- Published a short paper and presented the work at SIGIR'23.
- Built CIS multi-stage pipeline and won 2nd/7th Places in manual/automatic sessions in CAsT of Text REtrieval Conference 2022 (TREC'22).

Interactive Search: Condition-aware Retrieval for Personalized Search.

- Developed retrieval-enhanced clarification question generation (CQG).
- Incorporated CQG with open-domain conversational QA models for mixed-initiative response generation.
- Designed statement-aware retrievers with LLM-augmented personalized statements, and fine-tuned a statement-aware response generator for interactive search.

Retrieval-enhanced Applications: A Multi-stage Pipeline for Financial Analysis

- Defined a new signal highlighting task with a retrieval-enhanced pipeline for financial analysis; the proposed semi-supervised highlighting model outperforms the baselines by 12%.
- Published long papers and presented the work at EACL'23 and ACL'23.

Product Retrieval: Improving Lexical Product Retrieval with Product Content Reformulation

- Developed hybrid retrieval for TREC'23 Product Search Track and outperforms baselines by 15%.
- Improved lexical product search with text-to-text and image-to-text product contents reformulation.
- Designed a multimodal representation learning framework with text-image alignment and contrastive learning.

Cross-lingual Information Retrieval (CLIR): Improving Passage Re-ranking with Cross-lingual Query

- Constructed T5 passage re-ranking model with bilingual query for NeuCLIR track of TREC'23.
- Built CLIR systems and won 2nd, 3rd, 3rd Place (in 12 teams) in Chinese, Russian, Persian in NeuCLIR.

TEACHING EXPERIENCE

Management Information Systems Department, National Chengchi University

Teaching Assistant – Decision Science (master's-level course)

- Offered small group supervision on machine learning, statistics, and R/Python programming.

Taipei, Taiwan
Sep. 2020 - Feb. 2021

PUBLICATIONS

- **Jia-Huei Ju**, Sheng-Chieh Lin, Ming-Feng Tsai, and Chuan-Ju Wang. 2023. Improving Conversational Passage Re-ranking with View Ensemble. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'23)*, pages 2077–2081.
- **Jia-Huei Ju**, Yu-Shiang Huang, Cheng-Wei Lin, Che Lin, and Chuan-Ju Wang. 2023. A Compare-and-contrast Multistage Pipeline for Uncovering Financial Signals in Financial Reports. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL'23) (Volume 1: Long Papers)*, pages 14307–14321.
- **Jia-Huei Ju**^{*}, Ta-Wei Huang^{*}, Yu-Shiang Huang, Cheng-Wei Lin, Yi-Shyuan Chiang, and Chuan-Ju Wang. 2023. FISH: A Financial Interactive System for Signal Highlighting. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL'23): System Demonstrations*, pages 50–56. (* indicates equal contributions).
- **Jia-Huei Ju**, Jheng-Hong Yang, and Chuan-Ju Wang. 2021. Text-to-Text Multi-view Learning for Passage Re-ranking. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'21)*, pages 1803–1807.

SKILLS

- Languages: Mandarin (native), English (fluent)
- Programming: Python, R, Java, C++
- Machine learning: Hugging Face, PyTorch, JAX, TensorFlow
- Other: Vim, Linux, Google Cloud Platform & Cloud TPU

ACADEMIA-INDUSTRY COLLABORATION PROJECTS

- App Banner Recommendation Methods Based on Customers' Historical Behaviors Sep. 2022 - Jul. 2023
- Collaborated with AI R&D Center, E.SUN Commercial Bank.
 - Led 6 members to develop recommendation methods for selecting mobile banner ads to display.
 - Developed clustered CF methods and boosted baseline by 12% and reduced 50% of inference time.
- Analyzing Customer Characteristics using E-invoice Purchasing History May 2022 - Oct. 2022
- Collaborated with eCloud Mobile Corporation.
 - Designed a product name rewriting system for over 100,000 products in Chinese; system could identify same products with different names, which further benefited downstream customer analysis.
- Redesigning System Pipelines for Mutual Fund Recommendation Oct. 2021 - Sep. 2022
- Collaborated with AI R&D Center, E.SUN Commercial Bank.
 - Led 4 members to redesign system pipeline for mutual funds recommendation.
 - Constructed graph convolutional network RecSys and boosted original system's accuracy by 40%.
 - Designed content-based neural network recommender systems for solving cold-start users' issues.
- Estimating the Effects of Cross-Product Promotions on Sales Demand Dec. 2019 - Dec. 2020
- Collaborated with Industrial Technology Research Institute
 - Designed optimization algorithms for time-series estimation to capture cross-product promotion effects.
 - Conducted Monte-Carlo simulation experiments for empirical evaluation.