Shengyao Zhuang

T: (+61) 416190621 | shengyaozhuang@gmail.com | http://arvinzhuang.github.io/

PERSONAL SUMMARY

I am a postdoctoral researcher at CSIRO's Australian e-Health Research Centre, where I focus on developing large language model-based search engine systems in the medical domain. Before joining CSIRO, I was a Ph.D. student at the ielab, EECS, The University of Queensland, Australia, supervised by Professor Guido Zuccon. My primary research interests lie in information retrieval, large language model-based neural rankers, and NLP in general.

EMPLOYMENT HISTORY

The University of Queensland, Brisbane   **Dec 2023 – Current**
- Adjunct Lecturer

CSIRO, Australian e-Health Research Centre, Brisbane   **Mar 2023 – Current**
- Postdoctoral Research Fellow
- Search and Natural Language Process for Precision Medicine.

Microsoft, Software Technology Center Asia, Beijing   **May 2022 – Nov 2022**
- Research Intern
- Developing multimodal and multilingual entity retrieval models.

The University of Queensland, Brisbane   **Jul 2019 – May 2022**
- Tutor
- Tutoring of the INFS7410 (Information Retrieval and Web Search) course. Preparing and delivering course materials, assignments, and exam marking.

The University of Queensland, Brisbane   **Apr 2019 - July 2019**
- Casual Research Assistant
- Assist with the research and development of methods of information retrieval and evaluation based on online learning to rank.

EDUCATION

Ph.D. in Computer Science   **July 2019 – Dec 2023**
- University of Queensland, Brisbane
- Research topic: Teaching Pre-trained Language Models to Rank Effectively, Efficiently and Robustly
- Supervisor: Professor Guido Zuccon

Master of Information Technology   **Feb 2017 - Dec 2018**
- University of Queensland, Brisbane
- GPA: 6.3/7

Bachelor of Electrical Engineering   **Sep 2012 - Jul 2016**
- Chongqing University of Science and Technology, Chongqing China
- GPA: 79.64/100

AWARDS

2021 ICTIR: Best Student Paper Award
- Effective and Privacy-preserving Federated Online Learning to Rank
- https://sigir.org/sigir2021/awards/

2021 SIGIR Top10 Authors: Rank #3 (unofficial)
- Based on normalized full paper count.
- https://groups.cs.umass.edu/zamani/2021/04/29/sigir-2021-stats/

2021 Huawei DIGIX Global AI Challenge: CHAMPION
- Search rankings in multimodal and multilingual contexts.
- As the team leader, US$30,000 prize.
- https://developer.huawei.com/consumer/en/activity/digixActivity/Winnerlist

2022 SIGIR: Best Paper Honorable Mention
- Reduce, Reuse, Recycle: Green Information Retrieval Research
- https://sigir.org/sigir2022/program/best-paper-awards/

2022 ADCS: Best Paper Award
- Robustness of Neural Rankers to Typos: A Comparative Study
- http://adcs-conference.org/2022/index.html

PUBLICATIONS

PromptReps: Prompting Large Language Models to Generate Dense and Sparse Representations for Zero-Shot Document Retrieval
- **Shengyao Zhuang**, Xueguang Ma, Jimmy Lin, Guido Zuccon
- 2024, Arxiv (Under review)

Understanding and Mitigating the Threat of Vec2Text to Dense Retrieval Systems
- **Shengyao Zhuang**, Bevan Koopman, Xiaoran Chu, Guido Zuccon
- 2024, Arxiv (Under review)

ReSLLM: Large Language Models are Strong Resource Selectors for Federated Search
- Shuai Wang, **Shengyao Zhuang**, Bevan Koopman, Guido Zuccon
- 2024, Arxiv (under review)

A Setwise Approach for Effective and Highly Efficient Zero-shot Ranking with Large Language Models
- **Shengyao Zhuang**, Honglei Zhuang, Bevan Koopman, Guido Zuccon
- 2024 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2024)

Leveraging LLMs for Unsupervised Dense Retriever Ranking
- Ekaterina Khramtsova, **Shengyao Zhuang**, Mahsa Baktashmotlagh, Guido Zuccon
- Joint first author
- 2024 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2024)

Dense Retrieval with Continuous Explicit Feedback for Systematic Review Screening Prioritisation
- Xingyu Miao, **Shengyao Zhuang**, Bevan Koopman, Guido Zuccon
- 2024 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2024)

Large Language Models for Stemming: Promises, Pitfalls and Failures
- Shuai Wang, **Shengyao Zhuang**, Guido Zuccon
- 2024 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2024)

Stemming with Large Language Models: Promises, Pitfalls and Failures
- Shuai Wang, **Shengyao** Zhuang, Guido Zuccon
- 2024 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2024)

Fed4RAG: A New Test Collection for Federated Search Augmented Generation
- Shuai Wang, Ekaterina Khramtsova, **Shengyao Zhuang**, Guido Zuccon
- 2024 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2024)

Zero-shot Generative Large Language Models for Systematic Review Screening Automation
- Shuai Wang, Harrisen Scells, **Shengyao Zhuang**, Martin Potthast, Bevan Koopman, Guido Zuccon
- 2024, European Conference on Information Retrieval: Findings (ECIR2024-findings)

TeachingPpre-trained Language Models to Rank Effectively, Efficiently, and Robustly
- **Shengyao Zhuang**
- 2023, PhD Thesis

Open-source Large Language Models are Strong Zero-shot Query Likelihood Models for Document Ranking
- **Shengyao Zhuang**, Bing Liu, Bevan Koopman, Guido Zuccon
- 2023, Conference on Empirical Methods in Natural Language Processing: Findings (EMNLP2023-findings)

Typos-aware Bottlenecked Pre-Training for Robust Dense Retrieval
- **Shengyao Zhuang**, Linjun Shou, Jian Pei, Ming Gong, Houxing Ren, Guido Zuccon, Daxin Jiang
- 2023 ACM SIGIR Conference on Information Retrieval in the Asia Pacific (SIGIR-AP2023)

Selecting which Dense Retriever to use for Zero-Shot Search
- Ekaterina Khramtsova, **Shengyao Zhuang**, Mahsa Baktashmotlagh, Xi Wang, Guido Zuccon
- 2023 ACM SIGIR Conference on Information Retrieval in the Asia Pacific (SIGIR-AP2023)

Beyond CO2 Emissions: The Overlooked Impact of Water Consumption of Information Retrieval Models
- Guido Zuccon, Harrisen Scells, **Shengyao Zhuang**
- 2023 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR2023)

Exploring the Representation Power of SPLADE Models
- Joel Mackenziem**, Shengyao Zhuang**, Guido Zuccon
- Joint first author
- 2023 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR2023)

Augmenting Passage Representations with Query Generation for Enhanced Cross-Lingual Dense Retrieval
- **Shengyao Zhuang**, Linjun Shou, Guido Zuccon

- 2023 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2023)

Bridging the Gap Between Indexing and Retrieval for Differentiable Search Index with Query Generation

- **Shengyao Zhuang**, Houxing Ren, Linjun Shou, Jian Pei, Ming Gong, Guido Zuccon, Daxin Jiang
- 2023 The First Workshop on Generative Information Retrieval at SIGIR2023

Implicit Feedback for Dense Passage Retrieval: A Counterfactual Approach

- **Shengyao Zhuang**, Hang Li, Guido Zuccon
- 2022 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2022)

CharacterBERT and Self-Teaching for Improving the Robustness of Dense Retrievers on Queries with Typos

- **Shengyao Zhuang**, Guido Zuccon
- 2022 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2022)

Reduce, Reuse, Recycle: Green Information Retrieval Research

- Harrisen Scells**, Shengyao Zhuang**, Guido Zuccon
- 2022 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2022)

Asyncval: A Toolkit for Asynchronously Validation Dense Retriever Checkpoints during Training

- **Shengyao Zhuang**, Guido Zuccon
- 2022 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2022)

To Interpolate or not to Interpolate: PRF, Dense and Sparse Retrievers

- Hang Li, Shuai Wang, **Shengyao Zhuang**, Ahmed Mourad, Xueguang Ma, Jimmy Lin, Guido Zuccon
- 2022 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2022)

Fast Passage Re-ranking with Contextualized Exact Term Matching and Efficient Passage Expansion

- **Shengyao Zhuang**, Guido Zuccon
- 2022, 1st Workshop on Reaching Efficiency in Neural Information Retrieval (ReNeuIR)

Reinforcement Online Learning to Rank with Unbiased Reward Shaping

- **Shengyao Zhuang**, Zhihao Qiao, Guido Zuccon
- 2022 Information Retrieval Journal (IRJ)

Pseudo Relevance Feedback with Deep Language Models and Dense Retrievers: Successes and Pitfalls

- Hang Li, Ahmed Mourad, **Shengyao Zhuang**, Bevan Koopman, Guido Zuccon
- 2022, ACM Transactions on Information Systems (TOIS)

Improving Query Representations for Dense Retrieval with Pseudo Relevance Feedback: A Reproducibility Study

- Hang Li**, Shengyao Zhuang**, Ahmed Mourad, Xueguang Ma, Jimmy Lin, Guido Zuccon
- 2022 European Conference on Information Retrieval (ECIR2022)

Dealing with Typos for BERT-based Passage Retrieval and Ranking

- **Shengyao Zhuang**, Guido Zuccon
- 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP2021)

TILDE: Term Independent Likelihood model for Passage Re-ranking

- **Shengyao Zhuang**, Guido Zuccon
- 2021 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2021)

How do Online Learning to Rank Methods Adapt to Change of Intents?
- **Shengyao Zhuang**, Guido Zuccon
- 2021 ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2021)

Effective and Privacy-preserving Federated Online Learning to Rank
- Shuyi Wang, Bing Liu, **Shengyao Zhuang**, Guido Zuccon
- 2021 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR2021)

BERT-based Dense Retrievers Require Interpolation with BM25 for Effective Passage Retrieval
- Shuai Wang, **Shengyao Zhuang**, Guido Zuccon
- 2021 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR2021)

Deep Query Likelihood Model for Information Retrieval
- **Shengyao Zhuang**, Hang Li, Guido Zuccon
- 2021 European Conference on Information Retrieval (ECIR2021)

Federated Online Learning to Rank with Evolution Strategies: A Reproducibility Study
- Shuyi Wang, **Shengyao Zhuang**, Guido Zuccon
- 2021 European Conference on Information Retrieval (ECIR2021)

Counterfactual Online Learning to Rank
- **Shengyao Zhuang**, Guido Zuccon
- 2020 European Conference on Information Retrieval (ECIR2020)

REFERENCES

Name: Prof. Guido Zuccon
Organisation: UQ EECS
Position: Professor
Email: g.zuccon@uq.edu.au

Name: Dr. Bevan Koopman
Organisation: CSIRO AEHRC
Position: Senior Research Scientist
Email: bevan.koopman@csiro.au