GAO, YINGQIANG

Natural Language Processing · Large Language Models · Information Retrieval

Institute of Neuroinformatics | Winterthurerstrasse 190, 8057 Zurich, Switzerland

yingqiang.gao@ini.ethz.ch | https://charizardacademy.github.io | linkedin.com/in/yingqiang-gao/ | +41 78 670 7470

Education

ETH Zürich *Zürich, Switzerland*

Doctor of Science in Natural Language Processing *August 2020 - present*

- Institute of Neuroinformatics, Department of Information Technology and Electrical Engineering (D-ITET)
- Research topics in Argument Mining/Generation, Scientific Inference using Large Language Models, Information Retrieval for Massive Scientific Documents, Multilingual Neural Machine Translation
- Supervisor: Prof. Dr. Richard Hahnloser (D-ITET, ETHZ) and Dr. James Henderson (Idiap, EPFL)

Technische Universität München *München, Germany*

Master of Science in Computer Science *October 2017 - April 2020*

- Majored in Robotics, Cognition, Intelligence, School of Computation, Information and Technology
- Project experience in Sentiment Analysis, Biomedical Image Processing, Embedded Systems, Reinforcement Learning
- Supervisor: Prof. Dr. Georg Groh (Department of Computer Science, TUM)

Shanghai Tongji University *Shanghai, China*

Bachelor of Engineering in Mechatronics Engineering *September 2013 - August 2017*

- Majored in Mechatronics Engineering and Automation Technology, Faculty of Mechanical Engineering
- Project experience in additive manufacturing (3D printers and NC machines), automated control systems
- Supervisor: Prof. Dr. Nan Xie

Project Experience

Endoc: Scientific Writing Platform with Joint Document Retrieval and Text Generation

Description: A project that develops a modularized scientific writing platform with claim generation and fact verification for scientific documents. Endoc can jointly recommend relevant scientific documents, extract highlights, generate citations and claims, and retrieve the most relevant content.
Keywords: Document Retrieval/Processing, Natural Language Understanding/Generation, Large Language Models
Skills: Database (MongoDB, SQLite), Front-end (ReactJS, Streamlit, MUI), Back-end (Flask, GraphQL, ElasticSearch)
Page: `https://se-staging.ee.ethz.ch/`

Neurorobotics Robot Designer (The Human Brain Project)

Description: A Blender plugin to design simulation models of robots, biomimetic robots and musculoskeletal models faster and easier with a graphical user interface. Models can be exported in SDFormat and .osim format and are compatible with Gazebo and OpenSim.
Keywords: Neurorobotics, 3D Modeling/Simulation, Plugin Development
Skills: Simulation (Blender, CATIA), Programming (Python, C/C++)
Page: `https://github.com/HBPNeurorobotics/BlenderRobotDesigner`

Research Publications

Published

Anna Kiepura, **Yingqiang Gao**, Jessica Lam, Nianlong Gu, Richard H.R. Hahnloser. CODI 2024. *SciPara: A New Benchmark Dataset for Investigating Paragraph Discourse Structure in Scientific Papers*. In Proceedings of the 5th Workshop on Computational Approaches to Discourse (CODI'24), St.Julians, Republic of Malta. Association of Computational Linguistics.

Yingqiang Gao, Nianlong Gu, Jessica Lam, James Henderson, Richard H.R. Hahnloser. EACL 2024. *Evaluating Unsupervised Argument Aligners via Generation of Conclusions for Structured Scientific Abstracts*. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL'24), St.Julians, Republic of Malta. Association of Computational Linguistics.

Yingqiang Gao, Jessica Lam, Nianlong Gu, Richard H.R. Hahnloser. EMNLP 2023. *GreedyCAS: Unsupervised Scientific Abstract Segmentation with Normalized Mutual Information*. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP'23), Singapore, pages 6093-6108, Republic of Singapore. Association for Computational Linguistics.

Nianlong Gu, **Yingqiang Gao**, Richard H.R. Hahnloser. Workshop on Document Intelligence and Understanding (DocIU), co-located with CIKM 2023. *MemSum-DQA: Adapting An Efficient Long Document Extractive Summarizer for Document Question Answering*. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM'23), Birmingham, United Kingdom. Association for Computing Machinery.

Yingqiang Gao, Nianlong Gu, Jessica Lam, Richard H.R. Hahnloser. The 9th Workshop on Argument Mining, co-located with COLING 2022. *Do Discourse Indicators Reflect the Main Arguments in Scientific Papers?* In Proceedings of the 9th Workshop on Argument Mining, pages 34–50, Gyeongju, Republic of Korea. International Conference on Computational Linguistics.

Nianlong Gu, **Yingqiang Gao**, Richard H.R. Hahnloser. ECIR 2022. *Local Citation Recommendation with Hierarchical-Attention Text Encoder and SciBERT-Based Reranking*. In Proceedings of the 44th European Conference on Information Retrieval (ECIR'22), page 274-288, Stavanger, Norway. Advances in Information Retrieval.

Yingqiang Gao, Nikola I. Nikolov, Yuhuang Hu, and Richard H.R. Hahnloser. ACL 2020. *Character-Level Translation with Self-attention*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL'20), pages 1591–1604, Online. Association for Computational Linguistics.

Yingqiang Gao, Nan Xie, Kai Hu, Ying Zhu, Liang Wang. *An Optimized Clustering Approach Using Simulated Annealing Algorithm with HMM Coordination for Rolling Elements Bearings' Diagnosis*. Journal of Failure Analysis and Prevention 17, 602–619 (2017). Springer.

In Review

Yingqiang Gao, Jonathan Prada, Nianlong Gu, Jessica Lam, Richard H.R. Hahnloser. ACL 2024. *ENDOC: A Modularized Platform with Claim Generation and Fact Verification for Scientific Documents*. Paper under review.

Jessica Lam, Ximeng Cui, Shu Zhu, **Yingqiang Gao**, Nianlong Gu, Richard H.R. Hahnloser. NAACL 2024. *Exploring Metadata Matching for Reference Linking*. Paper under review.

Eren Öner, **Yingqiang Gao**, Jessica Lam, Nianlong Gu, Richard Hahnloser. SIGIR 2024. *Linear Narrative Flow Detection for Scientific Text*. Paper under review.

Awards

2023 **Top 1 Winner**, CIKM 2023, Workshop on Document Intelligence and Understanding (DocIU)

2014-2016 **National Scholarship of Outstanding Students**, Tongji University

2015 **Top 1 Winner**, Industry 4.0 International Student Competition of 3D printer Designing

2015 **Third Prize Winner**, National Mathematical Modeling Contest for Undergraduate Students

2014 **Best Student Referee**, The Soccer Association of Tongji University

Professional Experience

Language Abilities

German **Advanced communication (C1)**, Test Deutsch als Fremdsprache (TestDAF)

English **Advanced communication (C1)**, International English Language Testing System (IELTS)

Chinese **Native speaker**, National Proficiency Test of Putonghua (Mandarin) Certificate (Level A)

Technical Competencies

Python ★★★★★, PyTorch, TensorFlow, CUDA, Scikit-learn, Hugging Face, NLTK, Scipy, SpaCy

C/C++ ★★★★☆, MPI, OpenCV, CTranslate

JavaScript ★★★★★, ReactJS, Docker, MongoDB, SQLite, Streamlit, Flask, GraphQL, MUI

Linux ★★★★★, shell command, bash, git

Cloud ★★★★★, Google Cloud, DataFlow, Heroku

Professional Memberships

ETHZ **ETH AI Center**, Associated Doctoral Researcher

SwissNLP **Swiss Association of Natural Language Processing**, Academic Member

ETHZ **The Association of Scientific Staff at ETH (AVETH)**, Board Member & Politics Team Member

Endocite **The Endocite association**, Co-Founder & Board Member

ACL **Association of Computational Linguistics**, Student Member

Suyu **Zürich Suyu Chinese Ochestra (苏黎世苏遇民乐团)**, Founder & First Flutist

References

ETHZ **Prof. Dr. Richard Hahnloser**, rich@ini.ethz.ch

EPFL **Dr. James Henderson**, james.henderson@idiap.ch

UNIL **Dr. Aris Xanthos**, aris.xanthos@unil.ch

TUM **Prof. Dr. Georg Groh**, grohg@in.tum.de

Tongji **Prof. Dr. Nan Xie**, xienan115@tongji.edu.cn