# GAO, YINGQIANG · 高英强

Natural Language Processing · Large Language Models · Information Retrieval

*Institute of Neuroinformatics | Winterthurerstrasse 190, 8057 Zurich, Switzerland*

yingqiang.gao@ini.ethz.ch | charizardacademy.github.io | linkedin.com/in/yingqiang-gao/


## Education

**ETH Zürich** *Zürich, Switzerland*

Doctor of Science *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 *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 (上海同济大学)** *Shanghai, China*

Bachelor of 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 recommends relevant scientific documents, extract highlights, generates citations and claims, and retrieves 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`

## Publications

### Published

**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), pages XX–XX, Singapore, 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. NAACL 2024. *ENDOC: A Modularized Platform with Claim Generation and Fact Verification for Scientific Documents*. Paper under review.

**Yingqiang Gao**, Nianlong Gu, Jessica Lam, James Henderson, Richard H.R. Hahnloser. EACL 2023. *Evaluating Unsupervised Argument Aligners via Generation of Conclusions for Structured Scientific Abstracts*. Paper under review.

Anna Kiepura, **Yingqiang Gao**, Jessica Lam, Nianlong Gu, Richard H.R. Hahnloser. AAAI 2024. SciPara: A New Benchmark Dataset for Investigating Paragraph Discourse Structure in Scientific Papers. 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

## Academical Experience

| | |
|---|---|
| 2019 | **Visiting Master Student**, Institute of Neuroinformatics, Universität Zürich and ETH Zürich. Master thesis on topic character-level multilingual neural machine translation |
| 2019 | **Graduate Research Assistant**, Chair of Social Computing, TUM. Semester project on sentence-level aspect-based sentiment analysis using deep learning approach |
| 2018-2019 | **Graduate Research Assistant**, Chair of Robotics, Artificial Intelligence and Embedded Systems, TUM. The Human Brain Project (HBP), in charge of developing the neurorobotic platform (NRP) toolkit for Blender |
| 2016-2017 | **Undergraduate Research Assistant**, Helmholz-Zentrum Dresden-Rossendorf, Institute of Fluid Dynamics. DRESDYN project, bachelor thesis on power measuring and stream visualization for precession-driven dynamo experiments |
| 2014-2016 | **Undergraduate Research Assistant**, Numerical Control and Additive Manufacturing Lab, Tongji University. Construction of 3D printers and computerized numerical control turning lathe machines |

## Teaching Experience

| | |
|---|---|
| Spring 2022 | **227-0395-00L Neural Systems (master course)**, Teaching Assistant, Department of Information Technology and Electrical Engineering, ETHZ. |
| Autumn 2019 | **IN2210 Tracking and Detection in Computer Vision (master course)**, Teaching Assistant, Chair of Computer Aided Medical Procedure and Augmented Reality, TUM. |

## 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 |
| JavaScript | ★★★★★, ReactJS, Docker, MongoDB, SQLite, Streamlit, Flask, GraphQL, MUI |
| Linux | ★★★★★, shell command, bash, git |
| C/C++ | ★★★★☆, MPI, OpenCV, CTranslate |

### Professional Memberships

| | |
|---|---|
| ETHZ | **ETH AI Center**, Associated Doctoral Researcher |
| SiwssNLP | **Swiss Association of Natural Language Processing**, Academic Member |
| ETHZ | **The Association of Scientific Staff at ETH (AVETH)**, Board Member & Politics Team |
| 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 |