

Homepage

# Heng Yang

J +44-7878711663 |  $\blacksquare$  hy345@exeter.ac.uk |  $\P$  +44-7878711663 A Homepage | G GitHub | Huggingface | G Google Scholar Department of Computer, University of Exeter, United Kingdom



Version

#### OBJECTIVE

Seeking an opportunity to apply my expertise in genomics modeling and language modeling to contribute to impactful research and push the boundaries of AI4Science.

#### RESEARCH EXPERIENCE

- **GFM/LLM**: Designed and developed genomic language models from scratch. Created a comprehensive benchmarking framework for genomic foundation models. Studied the LLM pipelines and applications.
- Sentiment Analysis: Create one of the most popular open-source sentiment analysis framework.

#### **EDUCATION**

University of Exeter Sep 2021 - Sep 2025 PhD in Genomic LM and LLM Exeter, UK

South China Normal University Sep 2018 - Jun 2021 Master of Sentiment Analysis Guangzhou, China

Sep 2014 - Jun 2018 Yangtze University Bachelor of Computer Science Jingzhou, China

## ✓ PERSONAL OPEN-SOURCE PROJECTS

OmniGenBench – 5k Installations First large-scale in-silico benchmarking framework for genomic foundation models

• PyABSA – 350k Installations, 1k GitHub stars Aspect-based sentiment analysis framework, serving commercials, developers and scholars Sept. 2024 - Present

GitHub

June 2020 - Present

GitHub

## COMMUNITY CONTRIBUTION STATISTICS

- GitHub: 1.5k stars,  $\approx 160$  commits and 60 PRs/year. I am a open-source lover since the beginning of my research. I am grateful to the developers whose projects helped me a lot. Therefore, I am committed to share all my open-source projects on GitHub with friendly MIT lenience.
- Huggingface: 8 Models and 10 Spaces with  $\approx 1,000$ k downloads and 10k access, respectively. Thanks to the Hungingface platform, I am glad to share all my pretrained and fine-tuned state-of-the-art sentiment analysis and genomic foundation models. e.g., deberta-v3-base-absa-v1.1 and OmniGenome-186M. Moreover, I have been to releasing demos for low-resource research topics, like RNA secondary structure prediction and RNA Design.
- PyPi: 8 python wheels with  $\approx 800$ k downloads. I have made efforts to simplify the workflows and pipelines by packing and releasing open-access Python wheels. My wheels have been widely used in by the community. These wheels can be easily distributed via PyPi and installed via the pip tool.

#### MAIN PUBLICATIONS

OmniGenBench: Automating Large-scale Benchmarking for Genomic Foundation Models Heng Yang, Jack Cole, Ke Li

ArXiv Preprint

OmniGenome: Aligning RNA Sequences with Secondary Structures in Genomic Foundation Models Heng Yang, Ke Li

MPRNA: Unleashing Multi-species RNA Foundation Model via Calibrated Secondary Structure Prediction Heng Yang, Ke Li EMNLP 2024

The Best Defense is Attack: Repairing Semantics in Textual Adversarial Examples Heng Yang, Ke Li

EMNLP 2024

- PlantRNAFM: An Interpretable RNA Foundation Model for Exploration Functional RNA Motifs in Plants Haopeng Yu#, **Heng Yang**#, et al. (Co-first Author) Nature Machine Intelligence 2024
- Modeling Aspect Sentiment Coherency via Local Sentiment Aggregation Heng Yang, Ke Li

EACL 2024

PyABSA: A Modularized Framework for Reproducible Aspect-based Sentiment Analysis Heng Yang, Chen Zhang, Ke Li

CIKM 2023

- InstOptima: Evolutionary Multi-objective Instruction Optimization via LLM-based Instruction Operators Heng Yang, Ke Li EMNLP 2023
- BoostAug: Boosting Text Augmentation via Hybrid Instance Filtering Framework Heng Yang, Ke Li

ACL 2023

DaNuoYi: Evolutionary Multi-Task Injection Testing on Web Application Firewalls Ke Li, **Heng Yang**, Willem Visser IEEE Trans. on Software Engineering 2023

## $\mathbf{Y}$ Awards

PhD Scholarship, Research Grant

University of Exeter, 2021-2025

Chinese National Scholarship, First-class Academic Scholarships

South China Normal University, 2018-2020

**Outstanding Bachelor Graduate** 

Yangtze University, 2018