Heng Yang

📞 +44-7878711663 | ✉ hy345@exeter.ac.uk | 💬 +44-7878711663
🏠 Homepage | ⌂ GitHub | 🌐 Huggingface | G Google Scholar
Department of Computer, University of Exeter, United Kingdom

My
Homepage

Latest
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**: Developed genomic language models from scratch. Created a comprehensive benchmarking framework for genomic foundation models. Studied the LLM pipelines.
- **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
- **Yangtze University** — *Sep 2014 - Jun 2018*
  Bachelor of Computer Science — Jingzhou, China

</> PERSONAL OPEN-SOURCE PROJECTS

- **OmniGenBench** – 5k Installations — *Sept. 2024 - Present*
  First large-scale in-silico benchmarking framework for genomic foundation models — GitHub
- **PyABSA** – 350k Installations, 1k GitHub stars — *June 2020 - Present*
  The most popular aspect-based sentiment analysis framework, serving commercials and scholars — GitHub

🦾 COMMUNITY CONTRIBUTION STATISTICS

- **GitHub:** 1.5k **stars**, 180 **followers**, ≈ 160 **commits** and 60 **PRs/year**. I have been an 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 sharing all my open source projects on GitHub with friendly MIT lenience.

- **Huggingface:** 150 **likes** and 15 **followers**, sharing 8 **Models** and 10 **Spaces** with ≈ 1,000k **downloads** and 10k **access, respectively**. Thanks to the Hungingface platform, I am glad to share all of my pre-trained 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: Uploading** 8 **python wheels with** ≈ 800k **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*

- **Bridging Sequence-Structure Alignment in RNA Foundation Models**
  Heng Yang, Ke Li — **AAAI 2025**

- **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**

🏆 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

— 怡然而止 —

Appendix

♥ PERSONAL STATEMENT

I am deeply passionate about researching pre-trained and large-scale models. My work spans genomics modeling, sentiment analysis, adversarial attacks, and the development of open-source tools. With strong intrinsic motivation, I excel at independently solving problems and am skilled at exploring and implementing new solutions. To date, every project I have led has delivered significant results. Furthermore, I am committed to continuous learning and broadening my knowledge to effectively tackle future challenges.

</> MAIN PROJECTS

- **PyABSA** *Installations: 350k | GitHub Stars: 1k*
  Abstract: PyABSA is a modular framework built on PyTorch designed to democratize research in aspect-based sentiment analysis (ABSA). It supports over 31+ models and 30+ datasets across various ABSA subtasks, including aspect term extraction, sentiment classification, and end-to-end ABSA. PyABSA incorporates state-of-the-art ABSA models like LSA, which is featured in the **Stanford University 2022 AI Index Report** (Page 83). As the most popular ABSA framework, it serves both academic and commercial users and is the **first-recommended ABSA tool by ChatGPT.**
  Key Concepts: Data Augmentation, Embedding, Self-attention, Transformer, BERT, Pretraining, Finetuning, Online Hub, Huggingface Demonstrations, Software Design

- **OmniGenBench** *Installations: 5k*
  Abstract: OmniGenBench is the first open-source benchmarking framework designed to address challenges in genomic foundation models (GFMs). It supports diverse GFM architectures and integrates 42 million genome sequences across 75 datasets to evaluate various genomic tasks, including RNA structure prediction and phenotype classification. By providing user-friendly tools for finetuning and deployment, OmniGenBench facilitates the democratization of GFM applications.
  Key Concepts: GFM Architectures, Pretraining Objective Design, Data Curation/Augmentation, Mixture-of-Experts, Scalable Software Design, AMP, Task Formulation and Implementation, LLM Concepts

🗎 MAIN PUBLICATIONS

- OmniGenBench: Automating Large-scale Benchmarking for Genomic Foundation Models
  Heng Yang, Jack Cole, Ke Li (ArXiv Preprint)
  TLDR: Introduced OmniGenBench, an open-source framework that automates large-scale benchmarking for genomic foundation models, integrating millions of genomic sequences across numerous tasks to standardize and democratize genomic model evaluation.

- Bridging Sequence-Structure Alignment in RNA Foundation Models
  Heng Yang, Ke Li (*AAAI 2025*)
  TLDR: Proposed OmniGenome, an RNA foundation model that aligns RNA sequences with secondary structures, enabling bidirectional mappings and achieving state-of-the-art performance in RNA design and structure prediction tasks.

- MPRNA: Unleashing Multi-species RNA Foundation Model via Calibrated Secondary Structure Prediction
  Heng Yang, Ke Li (*EMNLP 2024*)
  TLDR: Developed MP-RNA, a multi-species RNA foundation model incorporating calibrated secondary structure predictions, enhancing performance in genomic tasks across species.

- The Best Defense is Attack: Repairing Semantics in Textual Adversarial Examples
  Heng Yang, Ke Li (*EMNLP 2024*)
  TLDR: Introduced Rapid, a novel approach employing adversarial detection and attack strategies to repair semantics in textual adversarial examples, enhancing NLP model robustness.

- PlantRNAFM: An Interpretable RNA Foundation Model for Exploring Functional RNA Motifs in Plants
  Haopeng Yu#, **Heng Yang**#, et al. (Co-first Author, *Nature Machine Intelligence 2024*)
  TLDR: Presented PlantRNA-FM, an interpretable RNA foundation model tailored for plant genomes, facilitating the discovery of functional RNA motifs and regulatory elements in plants.

- [Modeling Aspect Sentiment Coherency via Local Sentiment Aggregation](#)
  Heng Yang, Ke Li (*EACL 2024*)
  TLDR: Proposed a novel local sentiment aggregation paradigm to model aspect sentiment coherency, achieving state-of-the-art performance in aspect-based sentiment classification.

- [PyABSA: A Modularized Framework for Reproducible Aspect-based Sentiment Analysis](#)
  Heng Yang, Chen Zhang, Ke Li (*CIKM 2023*)
  TLDR: Introduced PyABSA, a modular framework built on PyTorch for aspect-based sentiment analysis, supporting over 31 models and 30 datasets, enhancing reproducibility in sentiment research.

- [InstOptima: Evolutionary Multi-objective Instruction Optimization via LLM-based Instruction Operators](#)
  Heng Yang, Ke Li (*EMNLP 2023*)
  TLDR: Presented InstOptima, a framework leveraging evolutionary algorithms and large language model-based instruction operators for optimizing multi-objective instructions, improving instruction clarity and task performance.

- [BoostAug: Boosting Text Augmentation via Hybrid Instance Filtering Framework](#)
  Heng Yang, Ke Li (*ACL 2023*)
  TLDR: Developed BoostAug, a hybrid instance filtering framework that enhances text augmentation techniques by maintaining feature space similarity with natural datasets, improving performance in NLP tasks.

- [DaNuoYi: Evolutionary Multi-Task Injection Testing on Web Application Firewalls](#)
  Ke Li, **Heng Yang**, Willem Visser (*IEEE Trans. on Software Engineering 2023*)
  TLDR: Introduced DaNuoYi, an evolutionary multi-task injection testing tool for web application firewalls, enhancing security testing by simulating diverse attack vectors to improve firewall robustness.