Ying Li

Mobile: (+86) 18311010597 Email: yingliclaire@pku.edu.cn

Summary

I obtained my Ph.D. and bachelor's degrees from Peking University and Nanjing University, respectively. I have extensive research experience in high-performance computing (HPC) and deep learning. I have published papers in prestigious journals and conferences such as EMNLP, Physical Review Letters, Advanced Functional Materials, and Physical Review Applied, with a total citation of 766 and an H-index of 15. My current research interests are primarily focused on large language models (LLMs) and LLM for math.

Education and Awards

Peking University, PhD, Computational Physics 2019.10 – 2024.07

- **China International College Students' Innovation Competition**, Bronze Award of Beijing Province, Intelligent Agent for Mental Health Monitoring and Emotional Companionship Based on LLM.
- **Amazon DeepRacer College Challenge**, second place of Peking University, representing Peking University in the national finals.

Peking University, PhD, Physics 2016.09 - 2019.10

Nanjing University, Bachelor in Physics 2012.09 - 2016.06

- **Mathematical Contest in Modeling**, Meritorious Winner
- **National Mathematical Modeling Competition**, First Prize of Jiangsu Province
- **Outstanding Graduate of Nanjing University**

Chongqing Nankai Middle School 2009.09 - 2012.06

Research Experience

Tsinghua University, YMSC and BIMSA 2025.03 - Present

- **LLM for Math Proof**
  Integrating LLMs with formal languages to achieve automated mathematical theorem proving through methods such as fine-tuning, reinforcement learning, prompt engineering, and data engineering.
- **Agent Based Modeling (ABM)**
  Introducing multi-agent reinforcement learning (MARL) algorithms into social behavior studies through LLM-based ABM, with a case study focusing on finite-horizon Markov games.
- **Multimodal Reasoning**
  Integrating multimodal data such as text, images, and tables to enhance the reasoning capabilities of large models, with applications in mathematical reasoning and scientific database construction.

Peking University, School of Computer Science 2024.06 – 2025.03

- **Distributed training and inference acceleration of LLMs**
  Data parallelism, tensor parallelism, and pipeline parallelism of LLMs. Accelerate inference and training through memory management. Adapt models to heterogeneous platforms (GPU + NPU).
- **Model ensembling and RAG (Retrieval-Augmented Generation)**
  Ensembling LLMs to achieve better generation than independent models. Building a RAG system with vector databases and embedding models, and optimizing system performance through reranker.

- **AI tutorials**
  I authored a series of AI tutorials for the **Ministry of Education's Large Model Public Service Platform** (Link). These tutorials cover neural network fundamentals, CNNs, LLM, and text-to-image generation.
- **Collaborative projects**
  I completed a series of projects related to LLMs as the Primary developer. The clients for these projects are the **Ministry of Education, China Unicom (Guangdong Branch), and a national research institute**.

Peking University, Laboratory of Computational Physics 2019.10 – 2024.07

- **Extracting Psychological indicators using NLP techniques**
  Fine-tune Transformer-based embedding models via an indirect supervised contrastive learning approach in Siamese network architecture to extract psychological indicators from historical corpora.
- **Predicting Hamiltonians with Graph Neural Networks (GNNs)**
  Treat the atoms and interactions between neighboring atoms as nodes and edges, respectively. Use GNN to update the embeddings of the nodes and edges to predict the elements of the periodic Hamiltonian.
- **Automated Density Functional Theory (DFT) workflow**
  Developed an automated DFT workflow that encompasses geometric structure processing, script generation, automatic task submission and error correction, and data post-processing.
- **Development and application of DFT methods**
  Develop DFT methods to address interface formation, defects, disorder, and other phenomena in interfaces and transport systems. Optimize the design and process issues of nanoscale transistors by DFT methods.

Media Coverage

- Co-corresponding Author work: Yuqi Chen, Sixuan Li, Ying Li, et al., EMNLP main, 2024. Covered by DeepTech and MIT Technology Review China.
- Co-first Author Work: Hong Li, Qiuhui Li, Ying Li (co-first author), et al., Advanced Functional Materials, 34(38): 2402474, 2024. Covered by Physics Department of Peking University.
- First Author Work: Ying Li, et al. Physical Review Applied, 20(6):064044, 2023. Covered by Fermi Technology.

Selected Publications

[1] Mohammad Atari, Yuqi Chen, **Ying Li**, *et al*. PsyEmbedding: Fine-tuning Pre-trained Transformer Models for Quantifying Psychological Similarity in Text. *In preparation*. 2025

[2] Sixuan Li, **Ying Li,** *et al.* AnaX: Anatomy-Aware Hybrid Network for ADHD fMRI Diagnosis. *submitted.* 2025

[3] Yuqi Chen, Sixuan Li, **Ying Li (co-corresponding author)**, *et al*. Surveying the dead minds: Historical-psychological text analysis with contextualized construct representation (CCR) for classical Chinese. *EMNLP main*, 2024.

[4] **Ying Li**, *et al*. Electrical contacts in monolayer $MoSi_2N_4$ transistor. *ACS Applied Materials & Interfaces*, 16(37): 49496–49507, 2024.

[5] **Ying Li**, *et al*. Monolayer WSi_2N_4: A promising channel material for sub-5-nm-gate homogeneous CMOS devices. *Physical Review Applied*, 20(6):064044, 2023.

[6] **Ying Li**, *et al*. Quantum transport simulation of sub-1-nm gate length monolayer MoS_2 transistors.

arXiv:2404.13801, Advanced Functional Materials (under review), 2024.

[7] **Ying Li**, *et al*. Ballistic MoS_2 transistors with ultra-high on-state current. *Science China Materials*, 67(10): 3083-3086, 2024.

[8] Hong Li, Qiuhui Li, **Ying Li (co-first author)**, *et al*. Recent experimental breakthroughs on 2D transistors: Approaching the theoretical limit. *Advanced Functional Materials*, 34(38): 2402474, 2024.

[9] **Ying Li**, *et al*. Review on quantum advantages of sampling problems. *Acta Physica Sinica*, 70(21): 210201, 2021.

[10] Benchuan Lin, Shuo Wang, Anqi Wang, **Ying Li**, *et al*. Electric control of fermi arc spin transport in individual topological semimetal nanowires. *Physical Review Letters*, 124(11):116802, 2020.

[11] Xinyue Yang, Shibo Fang, **Ying Li,** *et al.* Ab-Initio Quantum Transport Simulation of Sub-1 nm Gate Length Monolayer and Bilayer WSe_2 Transistors: Implications for Ultra-Scaled CMOS Technology. ACS Applied Nano Materials, 8(7): 3460-3470, 2025.

[12] Xingyuan Wang, Huazhou Chen, **Ying Li**, *et al.* Microscale vortex laser with controlled topological charge. Chinese Physics B, 25(12): 124211, 2016.

Google Scholar Link

Internship

Huawei Technologies Co., Ltd. 2012 Lab, Data Center 2021.01-2021.06

- **Quantum machine learning**
  Combination of quantum circuits and machine learning, including quantum convolutional neural networks, quantum-classical hybrid neural networks, and quantum variational circuits.
- **Classic simulator for random quantum circuits sampling**
  Calculate the sampling probability of the Sycamore circuit using tensor networks and reduce the actual simulation time by incorporating the number of read and write operations into the loss function.

China Merchants Securities Co., Ltd. Information and Communications Team 2020.08-2020.11

- Industry data analysis, research report writing, IPO prospectus review

Technology Stack

Coding

- Programming Languages: Python (proficient), bash (proficient), C++, C, Matlab, Mathematica
- Tools: git, docker
- Frameworks: PyTorch, HuggingFace (Transformers, Accelerate, Diffuser, Gradio, etc.), DeepSpeed, TensorFlow, vLLM, LangChain, Langgraph, MindIE, LLaMA-Factory
- HPC Platforms: HPC Platform of Tsinghua University; National Supercomputing Shenzhen Center; National Supercomputing Guizhou Center; HPC Platform of Peking University; In-house Server Cluster
- Computational software and packages: VASP, QuantumATK, Quantum ESPRESSO, SIESTA, Pymatgen, ASE, GPAW, Qiskit, Cirq, etc.

Operations & Maintenance Experience

- Administrator: Data Center for Laboratory of Computational Physics
- Executive AP (committee member) of Linux Club of Peking University (LCPU)

GitHub Link Gitee Link

Teaching Experience

For 8 universities including PKU, BJTU, etc: XSCOW Platform Training, Lecturer (Link) 2024
Peking University: Programming Language Series Lectures, Lecturer 2021
Peking University: Solid State Physics, Teaching Assistant 2017

Mentoring:

Fall 2017, Jia Yu, BS Physics (Peking University ’18)
Fall 2017, Jianwei Wang, BS Physics (Peking University ’18)

Talks

An Introduction to Large Language Models： Foundations and Applications 2025.03.28

- Invited by Shenzhen Institute for Quantum Science and Engineering of Southern University of Science and Technology

Automated Formalization and Theorem Proving via Large Language Models 2025.05.27

- Postdoc Workshop III 2025, Yau Mathematical Sciences Center, Tsinghua University

Activities

Peking University, AI for Science Salon 2023.10.28

- Planned and co-organized a salon themed “AI for Science”. Invited researchers from ByteDance Research, DP Technology, and Microsoft Research Asia to share their insights, with over 100 participants.