Kai Ling

ling59@purdue.edu • (765)586-8733 • LinkedIn • Website

PROFILE

Computer Science PhD candidate (Purdue) building large-scale deep-learning systems for molecular structure prediction — Transformer and diffusion models, distributed multi-GPU training and inference, and model distillation from open foundation-model stacks. Benchmarks rigorously against state-of-the-art (AlphaFold3, Boltz-2, Chai-1) and publishes. Comfortable owning the full ML lifecycle: data pipelines, training, evaluation, and production serving.

EDUCATION

Purdue University **West Lafayette, USA**

PhD (expected) Department of Computer Science Aug. 2019 – Present

Huazhong University of Science and Technology (HUST) **Wuhan, China**

Bachelor of Science in Computer Science and Technology Aug. 2015 – July 2019

GPA: 3.8/4.0

SELECTED WORK & RESEARCH EXPERIENCES

Research Assistant | Kihara Lab | Purdue **West Lafayette, USA**

Advisor: Daisuke Kihara, Professor, Purdue University

Lead | NuFold Multimer: RNA-RNA Complex Structure Prediction Jan. 2026 – Now

- Extended the lab's single-chain NuFold model to multi-chain inputs with a diffusion-based structure module and multi-sample conformational-ensemble generation — the first dedicated deep-learning predictor for RNA–RNA complexes.
- Built a large-scale model-distillation pipeline to overcome scarce ground truth: ran teacher-model inference (OpenFold3, Protenix) over ~20k RNA–RNA complexes with confidence-based filtering, after evaluating each stack under license / throughput / input-format constraints.
- Curated a multi-source training set (RNAInter, RISE, snoDB) with tiered confidence stratification; debugged cross-database label errors — strand-orientation, lncRNA/protein-coding misclassification, ID contamination — to produce clean supervision.
- Scaled training crops to ~1,000 tokens — beyond AlphaFold3's 768 — on substantially fewer GPUs by integrating NVIDIA cuEquivariance kernels and a memory-efficient diffusion module.
- Benchmarked SOTA predictors (AlphaFold3, Boltz-2) on RNA–RNA complexes and showed none reliably handle them, establishing the need for a dedicated model.

Lead | Distance-AF Multimer: Constraint-Guided Protein Complex Modeling Feb. 2024 – Sept. 2025

- Built a distributed multi-GPU optimization pipeline (PyTorch) for protein-complex structure prediction on 50k+ atom graphs.
- Adapted the lab's Distance-AF method to complexes as a per-target, test-time optimization that fits user-supplied distance constraints with no retraining — robust to noisy and partial constraints.
- Designed a coarse-to-fine, two-stage optimization that escapes the local minima from AlphaFold-Multimer's poor initialization, eliminating flying residues and steric clashes that otherwise stall convergence.
- Benchmarked across three datasets (27 hard multimeric targets, 8 peptide complexes, 3 large assemblies with 10+ chains); outperformed Chai-1, Protenix, and AlphaLink2 on Win/Tie/Loss, with up to ~60Å global-RMSD improvement over AFM on the hardest targets. First-author manuscript in preparation.

Designer & Sole Developer | DAQ-Refine: Cryo-EM Protein Model Refinement Service Aug. 2023 – Jan. 2024

- Deployed the lab's cryo-EM refinement method as a production web service (React + Flask, Mol* 3D viewer); open-access, used by labs worldwide at ~10–20 jobs/month, and part of the published EMSuite server.
- Re-architected from a usage-capped Colab notebook to a distributed backend, removing per-user limits and enabling large multimer jobs; runs and compares 3 refinement strategies and auto-selects the best.

CS577 | Natural Language Processing, Purdue University **West Lafayette, USA**

Lead — Image Captioning with Multi-Task Learning Jan. 2022 – Apr. 2022

- Built an end-to-end captioning pipeline (PyTorch) on MSCOCO (330K images): CNN encoders (ResNet50, EfficientNet-B0) with an attention-based LSTM decoder.
- Added multi-task auxiliary objectives and an object-attention alignment mechanism, improving BLEU-4 from 0.122 to 0.142 (+16.4%) over baseline.

Research Assistant | MSSN Lab | Purdue **West Lafayette, USA**

Advisor: Chunyi Peng, Professor, Purdue University

- Built mperf, a Django web platform for large-scale 5G data collection (sole developer, first version), and the lab's network-data visualization map — adopted as the ACM IMC 2018 poster.
- Ran large-scale 3GPP R17 protocol analysis with a lightweight formal method (Delta), finding 4 security vulnerabilities and building 12 proof-of-concept exploits.
- Led large-scale cellular measurement studies — a nationwide 5G study across major US carriers (mmWave + Sub-6 GHz), root-causing systemic underutilization and prototyping the "5GBoost" patch, plus the first city-scale cellular-UAV study (45 hr, 1,010 km) with deep RAN-layer trace analysis.

PUBLICATIONS(SELECTED)

First-author

- **Ling, K.** et al. Peptide–protein docking: from physics-based models to generative intelligence. *Chemical Communications*, 2026.
- **Ling, K.** et al. Distance-AF Multimer: A Multi-Step Approach for Protein Complex Structure Modeling with User-Defined Distance Constraints. *In preparation*.
- **Ling, K.*** et al. MMMap: Mobility Management Map of Global Carrier Networks. *ACM IMC 2018* (poster, *co-first).

Co-author

- Li, Y. et al. (incl. **Ling, K.**). Experience: A Five-Year Retrospective of MobileInsight. *ACM MobiCom 2021*.
- Deng, H., **Ling, K.** et al. Unveiling the Missed 4.5G Performance in the Wild. *ACM HotMobile 2020* (paper + poster).
- Cryo-EM EMSuite server / EM Webserver. *Biophysical Society Annual Meeting abstracts, 2025–2026.*

TEACHING AND LEADERSHIP

- **Teaching Assistant** — CS536 Computer Networking, Purdue University · Spring 2020 – Spring 2025. Led office hours and debugging sessions, authored assignments and exam questions, and graded across multiple cohorts — explaining complex networking systems and protocols to students.
- **Piano & Chorus**, HUST — 20+ concerts; reached national finals, China University Students Arts Festival.
- **Volunteers, HUST** — Volunteered for the Internet Innovation and Entrepreneurship Competition in 2016 in China.

AWARDS

1. Freshmen Scholarship (1 out of 120), HUST	2015
2. National Scholarship (2 out of 120), HUST	2017
3. Model Student of Academic Records (5 out of 120), HUST	2017
4. Awarded Outstanding Student (4 out of 120), HUST	2017

SKILLS

- **Programming Languages**: Proficient in Python, C++, Java, SQL, Bash;
- **ML / DL:** PyTorch, Transformers, diffusion models, distributed & multi-GPU training/inference, conformational-ensemble generation, model distillation; structure-prediction stacks (AlphaFold2/3, Boltz-2, OpenFold3, Protenix, NuFold); NVIDIA cuEquivariance GPU kernels;
- **LLM / GenAI:** Anthropic/OpenAI APIs, RAG, vector DBs (FAISS/pgvector), function-calling / agents, evaluation pipelines;
- **Full-stack:** React, Flask, Django, RESTful API design, Mol* / interactive 3D visualization;
- **Infra / DevOps:** Docker, Git, CI/CD, Linux, HPC / SLURM (distributed training & inference);
- **LLM / GenAI:** Anthropic/OpenAI APIs, RAG, vector DBs (FAISS/pgvector), function-calling / agents, eval pipelines;