

Ruochen Li

+49 1623382928 | [GitHub Page](#) | [Google Scholar](#) | liruochen8@gmail.com

Research Focus: Reliable Multimodal Learning & Alignment – Evaluation, Post-training, and Interpretability

EDUCATION

Oct. 2022 – Mar. 2026	M.Sc. in Computer Science, Technical University of Munich	GPA: 1.5 / 1.0, Distinction
Sep. 2017 – Jun. 2021	B.Sc. in Computer Science, Nanjing Forestry University	GPA: 87.6 / 100, Top 5%

SELECTED RESEARCH EXPERIENCE

Reliable LLM/VLM Evaluation & Post-training Alignment

1. ReEvalMed: Rethinking Medical Report Evaluation by Aligning Metrics with Real-World Clinical Judgment

R. Li*, J. Li*, B. Jian, K. Yuan, Y. Zhu **EMNLP 2025 Main** [\[paper\]](#)

- Built a fine-grained, clinician-validated meta-evaluation benchmark for assessing radiology report metrics (X-ray)
- Revealed clinical semantic misinterpretations and score-inflating evaluation practices in widely-used metrics
- Provided a framework for developing clinically reliable evaluation metrics and medical report generation models

2. Beyond Scalar Scores: Exploring LLM-based Metrics for Clinical Significance Evaluation in Radiology Reports

Q. Lu*, R. Li*, ..., D. Tao **Manuscript under review**

- Benchmarked LLM evaluators across model families, revealing consistent discrimination bias and robustness failures
- Designed a 4K synthetic dataset for training LLM-based evaluators on fine-grained clinical significance assessment
- Post-trained Qwen-8B and MedGemma-4B with SFT/DPO + LoRA, outperforming 32B medical LLMs in clinical alignment

3. SurgGoal: Rethinking Surgical Planning Evaluation via Goal-Satisfiability

R. Li*, K. Yuan*, ..., N. Navab **Preprint, under review** [\[arXiv\]](#)

- Proposed a paradigm shift in surgical planning evaluation: from sequence similarity to safety-critical goal-satisfiability
- Showed sequence similarity metrics can reward unsafe step orderings and penalize clinically valid alternative plans
- Diagnosed Video-LLM perception failures; structural (vs. semantic) knowledge better constrains surgical planning

LLM Interpretability & Representation Learning

4. Focus Directions Make Your Language Models Pay More Attention to Relevant Contexts

Y. Zhu, R. Li, D. Wang, D. Haehn, X. Liang **Preprint, under review** [\[arXiv\]](#)

- Identified distraction-related contextual heads in long-context LLMs and analyzed their role in task misalignment
- Proposed training-free focus directions in key/query activations that steer model attention toward task-relevant context
- Improved long-context task alignment by mitigating distraction from irrelevant contexts across multiple LLM families

5. Classification, Regression and Segmentation directly from k-Space in Cardiac MRI

R. Li*, J. Pan*, ..., D. Rueckert **MICCAI Workshop** [\[paper\]](#)

- Derived representation from UK Biobank k-space using Masked Autoencoders for classification/regression/segmentation
- Showed Vision Transformers better capture global frequency-domain dependencies than CNNs in complex-valued k-space
- Enabled robust prediction from undersampled, human-unreadable k-space data, competitive with image-input methods

ONGOING RESEARCH & ADDITIONAL PUBLICATIONS

6. Agentic Metric for Medical Generated Report Evaluation

Ongoing collaboration with Jean-Philippe Corbeil **Microsoft Healthcare** [\[LinkedIn\]](#)

- Developing a clinician-aligned LLM-agent metric for fine-grained, generalizable evaluation of medical report generation

7. Multimodal CT Radiomics–Clinical Model for Hepatic Steatosis Risk Prediction with Metabolomic Interpretation

Y. Xia, R. Li, ..., X. Wang **Frontiers in Nutrition** [\[paper\]](#)

- Built a multimodal model for PNAHS risk prediction, using metabolomics to interpret risk-associated imaging differences

8. Learning cascade attention for fine-grained image classification

Y. Zhu, R. Li, Y. Yang, N. Ye **Neural Networks** [\[paper\]](#)

- Developed a fully weakly supervised, parallel CNN architecture with spatial and cross-network attention mechanisms

9. Temporal Conditioning for Longitudinal Brain MRI Registration and Aging Analysis

B. Jian, ..., R. Li, D. Rueckert, B. Wiestler, C. Wachinger **IEEE TMI** [\[paper\]](#)

- Developed a learning-based framework for sparse-scan longitudinal brain MRI registration and brain-state forecasting

SKILLS & ACADEMIC SERVICE

Languages	English (C1), Mandarin (Native)
Programming	Python (PyTorch, Hugging Face Transformers, NumPy), C/C++, SQL, Bash, Java, HTML/PHP/Javascript
Reviewer	IEEE Transactions on Medical Imaging (TMI); EMA4MICCAI 2025 Workshop