黄梓洵

zixunh@wharton.upenn.edu | 个人主页 | GitHub | 谷歌学术 | LinkedIn | 微信：18952009927

研究兴趣

机器学习的理论基础，重点关注优化理论与标度律、大语言模型推理、后训练与智能体强化学习，以及人工智能应用。

教育经历

宾夕法尼亚大学沃顿商学院，统计学与数据科学博士 2026.08 – 至今

北京大学数学科学学院，统计学理学学士；应用数学拔尖学生培养计划；GPA：3.7/4.0 2022.09 – 2026.06

加州大学伯克利分校，访问学生；GPA：4.0/4.0 2025.01 – 2025.08

工作经历

腾讯混元——青云人才计划，研究实习生 2026.06 – 2026.08

论文发表

[1] Jia-Nan Wang*, **Zixun Huang***, Kairui Li*, Lei Wu. *Momentum in Large-Batch Training: Polyak Enlarges the Critical Batch Size, Nesterov Improves Data Efficiency.* **预印本**。[论文]

[2] **Zixun Huang***, Kishan Panaganti*, Haitao Mi, Leowei Liang. *FlowBalance: Verifier-Grounded Self-Improvement from On-Policy Reasoning Experience.* **预印本**。[论文] [GitHub] [博客]

[3] Zhongzhi Li*, Yucheng Shi*, Zongxia Li*, Ruhan Wang, Anhao Li, **Zixun Huang**, Junyao Yang, Lei Ke, Ninghao Liu, Haitao Mi, Leowei Liang. *Recursive Synthesis for Long-Horizon Terminal Tasks.* **预印本**。[论文] [Hugging Face] [项目主页]

[4] **Zixun Huang***, Jiayi Sheng*, Zeyu Zheng. *Variance-Aware Baselines and Adaptive Learning Rates for Reinforcement Learning with Verifiable Rewards.* **审稿中**。[论文]

[5] Yuhao Sun, Zekun Wu, **Zixun Huang**, Peijie Zhou. *TracingFlow: A Simulation-Free Trajectory Inference Framework Based on Second-Order Dynamics.* **审稿中**。[论文]

[6] Binghui Li*, Fengling Chen*, **Zixun Huang***, Lean Wang*, Lei Wu. *Functional Scaling Laws in Kernel Regression: Loss Dynamics and Learning Rate Schedules.* **NeurIPS 2025 Spotlight**。[论文] * 共同贡献。

科研经历

优化理论与标度律

▸ 北京大学，研究实习生，北京 2023.05 – 2026.08

导师：吴磊教授

- **大批量动量方法 [1]**：聚焦大批量训练中并行效率与数据效率的核心权衡，建立 Polyak 与 Nesterov 动量的统一标度理论：Polyak 通过加速信号学习扩大临界批量，使训练在提高并行度的同时保持最优数据标度率；Nesterov 则利用前瞻机制实现曲率自适应降噪，在大批量区间进一步提升数据效率。完整刻画批量大小相图，并通过数值实验验证稳定边界、临界批量与损失标度律。
- **函数型标度律 [6]**：基于幂律核回归建立单遍 SGD 的内在时间分析，将学习率调度的影响表述为卷积泛函，并推导可描述任意调度下完整损失轨迹的函数型标度律；进一步给出常数、指数衰减及 warmup–stable–decay 调度在数据受限与算力受限区域的显式关系，并通过核模拟和 0.1B–1B 参数语言模型预训练验证其拟合、预测与调度比较能力。

大语言模型推理、后训练与智能体强化学习

▸ 腾讯混元——青云人才计划，研究实习生，北京 2026.06 – 2026.08

导师：高级研究员 *Kishan Panaganti*

- **验证器引导的自我改进 [2]**：提出 FlowBalance，通过轨迹平衡将验证器优势与特权后见自引导相结合，学习完整回答上的归一化分布；在五个推理基准上，Qwen3-4B 与 Qwen3-8B 相较 GRPO 的平均表现分别提升 **1.95** 和 **2.12** 个百分点，同时显著增加成功解答之间的语义策略多样性。
- **长程智能体训练 [3]**：构建递归验证任务合成框架 RST：扩展已有任务的参考解法，同步重写指令与验证器，并在全新沙箱中执行通过后将其作为下一轮种子；15 轮共生成 **37,484 个任务**，单任务成本约 0.05 美元，参考解法中位长度由 67 行增至 374 行。基于合成任务采集 Qwen3.5 轨迹，通过监督微调与智能体 PPO 在三个长程终端基准上取得稳定提升。

▸ 加州大学伯克利分校，研究实习生，美国加州伯克利 2025.01 – 2025.08

导师：郑泽宇教授

- **方差感知策略优化 [4]**：针对带 KL 正则的可验证奖励强化学习（RLVR），证明策略梯度估计的无偏性，推导包含 KL 交叉协方差的精确方差及优化损失上界，并给出收敛性保证；据此设计梯度加权的方差最优基线与由策略梯度信噪比驱动的自适应学习率，形成 OBLR-PO。在 Qwen3-4B-Base 的四个推理基准上，平均 Pass@1 相较 GRPO 提升 **2.52** 个百分点。

部分荣誉与奖励

“挑战杯”竞赛特等奖	2026	北京大学未名学者	2026
鸿升奖学金	2024	全国大学生数学竞赛一等奖	2024
丘成桐大学生数学竞赛团体奖	2024	小米奖学金	2023
中国数学奥林匹克金牌	2021	中国数学奥林匹克银牌	2020

技能

编程： Python（NumPy、pandas、scikit-learn、PyTorch）、C++、MATLAB、LaTeX

语言： GRE V160/Q170；GRE 数学专项 970（第 95 百分位）；托福 101；CET-6 578