

机器学习与认知实验室

研究方向:

资源高效的机器智能：用更少的信息与算力，催生更强的智能

研究内容:

AI正在从“规模竞赛”走向“效率竞赛”。当前人工智能能力提升仍高度依赖更多数据、更大算力和持续反馈，但真实世界中的高质量数据、可靠反馈和计算资源往往有限且昂贵。**AI发展的下半场，不是无限堆叠资源，而是让模型和智能体用更少资源获得更强能力。**围绕这一问题，课题组形成“机理理解—方法设计—场景验证”的研究体系。

1. 原理：为什么少量信息也能学？为什么有时又会突然失效？研究模型与智能体如何利用有限样本、上下文和反馈实现泛化，以及何时失效；重点关注上下文学习、大模型训练动力学和低精度稳定性。
2. 方法：怎样少看样本、少做试错，也能快速学会新任务？利用少量样本、结构化上下文、稀疏反馈和历史经验，降低模型训练与智能体适应成本；研究快速泛化、模型适配、经验复用和测试时适应。
3. 应用：当每一次试错都很贵，AI还能不能可靠完成复杂任务？面向高成本、低监督的复杂真实任务，研究资源高效方法在分子科学、药物安全等场景中的应用，推动模型从预测走向自主决策与可靠执行。

代表性成果：

- 让大模型训练不再“突然崩掉”：揭示低精度 Flash Attention 训练失稳机制并提出稳定化方法，解决长期存在的 Loss Explosion 问题；相关机制被 Kimi 大模型训练采用（ICLR 2026 Oral）。
- 解释“大模型为什么看几个例子就会做”：揭示上下文学习中的泛化机制，为大模型利用少量示例快速适应提供理论解释；相关理论被 Berkeley David Culler 教授团队用于高效大模型服务研究（ICLR 2025）。
- 让多个大模型“学会交流”：提出语言模型网络 LMNet，通过模型间密集通信提升有限监督下的协同学习能力；被 Google DeepMind 评价为多模态模型与多智能体通信的代表性方法（ICML 2026）。
- 让AI预测“从未见过的新药”：提出 EmerGNN，利用生物学关系证据解决新药缺少历史标签时的相互作用预测与解释问题；获自然子刊、麻省理工科技评论专文评述（Nature Computational Science 2023）。



姚权铭，清华大学电子工程系副教授、IET Fellow。主要研究资源高效的机器智能，在小样本学习、大模型学习机理、智能体和科学智能等方向开展研究。发表高质量论文100余篇，Google Scholar引用超过1.6万次。担任IEEE TPAMI Action Editor、Neural Networks Senior Action Editor、Machine Learning Action Editor等，长期担任ICML、NeurIPS、ICLR等国际顶级会议领域主席。入选AAAI New Faculty Highlights、IJCAI Early Career Spotlight、IEEE Computing's Top 30，获INNS Aharon Katzir Young Investigator Award、蚂蚁InTech科技奖、华为火花奖等。主持政府及企业科研项目近30项，累计签约经费约2000万元，并与华为等企业保持长期科研合作。