

顶会观察

NeurIPS 2025

南京大学人工智能学院副教授 叶翰嘉

神经信息处理系统大会（Conference on Neural Information Processing Systems）作为机器学习领域的顶级盛会，持续保持 CCF-A 类会议的权威地位，在全球学术界与产业界拥有深远影响力。第 39 届 NeurIPS 大会创新采用“双会场+全虚拟并行”的举办模式，分别于 2025 年 11 月 30 日至 12 月 5 日在墨西哥城希尔顿 reforma 酒店、2025 年 12 月 2 日至 7 日在美国圣地亚哥会议中心设立线下会场，并同步开放全球虚拟参会通道。会议日程涵盖多元学术环节，包括专题 Tutorial、行业博览会、核心学术会议、海报展示、前沿研讨会等，其中墨西哥城会场设置 6 场线下 Tutorial 与 8 个主题 Workshop，圣地亚哥主会场则围绕强化学习、AI 安全、大模型优化等核心方向开展 3 天正式会议及 2 天深度 Workshop，通过直播联动实现学术内容互通，为全球约 2.6 万余名注册参会者打造沉浸式、跨地域的学术交流体验。

一、会议亮点

首创“双会场+全虚拟”并行模式：本届 NeurIPS 会议突破往届单一举办地限制，首次在墨西哥城与圣地亚哥设立双线下会场，搭配全球虚拟参会通道，在满足不同地区参会者的线下交流需求的同时，又通过直播联动技术实现双会场内容同步互通，兼顾学术交流的沉浸感与覆盖广度。

评审机制革新：针对投稿量激增 37.7%，会议延续并强化“Responsible Reviewing”相关举措：对未履行评审义务的作者可采取限制 rebuttal 期间获取评审意见等措施；对严重失职（如低质/敷衍）评审者，其合

著投稿可能被 desk-reject；明确 LLM 使用规范，要求作者披露相关使用情况。评审团队扩容至 2.2 万余人，通过优化校准流程维持 24.5% 录用率，同时强化保密机制，保障评审公正安全。

学术赛道拓展：新增立场论文赛道，聚焦 AI 伦理等宏观议题；引入期刊整合赛道，精选 34 篇顶级期刊成果展示，打通资源壁垒。数据集与基准测试赛道投稿量达 1995 篇，成为独立核心赛道，推动 AI 研究向数据质量与基准公平性深化。

二、录用情况

投稿与录用：大会共收到了 **21,575 篇**有效投稿，相较于 2024 年增长了约 **37.7%**，十年间投稿量更是从 2015 年的 1,838 篇增长超十倍，创下历史最高纪录，充分彰显了该领域全球研究热度的爆发式提升。经过多轮严格评审，最终接收 **5,290 篇**论文，录用率约为 **24.52%**，延续了近年“投稿量激增但录用率稳定在 24%-26% 区间”的趋势。在录用的论文中，仅有 **77 篇**被选为 Oral Presentations，**Oral 率仅为 0.36%**，不仅显著低于往年水平，也远低于同领域其他顶会，成为本届会议极具含金量的学术荣誉。此外，另有 **688 篇**论文被选为 Spotlight（重点海报），约占投稿总数的 **3.19%**，其余 4,525 篇录用论文以普通 Poster 形式展示，整体录用结构凸显了会议对学术成果质量的追求。

评审团队：为了处理海量投稿，会议组织了庞大的评审团，包括 21921 名技术审稿人、641 名伦理审稿人、1985 名领域主席和 240 名高级主席，同时优化评审校准流程，在投稿量翻倍的情况下仍维持稳定录用率，

确保质量门槛不降低。

中国力量：据第三方统计，NeurIPS 2025 接收的 5290 篇主赛道论文中，**中国及港澳台区域机构参与的论文占比接近 50%**，成为全球贡献量最高的区域。这一比例较 2021 年（约 35%-40%）显著提升，反映出中国在机器学习与 AI 领域的科研生产力已实现从**重要参与者到核心引领者**的转变。中国内地高校与企业是 NeurIPS 2025 中国力量的核心参与主体。高校方面，清华、北大等顶尖院校稳居机构贡献榜前列，其中清华大学以 138 篇录用论文位列**全球第二、高校第一**；港澳台高校持续稳定输出，合计贡献约 80 篇录用论文。企业方面，阿里巴巴以 146 篇录用论文位居**全球企业第三**，华为、腾讯、字节跳动等合计超 200 篇，覆盖大模型、图神经网络、数据集基准测试等多个关键方向，且产出多篇高质量论文与口头报告成果。

三、热门研究方向

根据投稿和录用分布，NeurIPS 2025 呈现出以下技术趋势：

架构革新聚焦效率优化：彻底告别盲目堆参的发展模式，转向“高效智能”的核心诉求。线性复杂度架构（如 YAAD、MONETA）表现亮眼，在多项基准测试中反超传统 Transformer，打破其长期垄断地位。同时，动态计算分配、200 万 token 超长上下文窗口等技术突破，既解决了超长序列处理难题，又实现了计算效率的线性增长，让模型性能与部署效率同步优化。

智能体与具身智能全面崛起：智能体相关研究成为主流，“长上下文、效率优化、智能体/具身、安全与评测”等成为高频主题。核心探索聚焦大规模智能体协作（解决数千智能体的通信过载与目标对齐问题）与零样本机器人控制（如通过 YouTube 视频训练“世界模型”，实现机器人无专门数据即可模仿人类动作），推动 AI 从“对话工具”向“可行动实体”演进，而多模态融合技术则为这一演进提供关键支撑。

研究导向转向数据与可信：AI 研究从“模型中心”逐渐转向“数据中心”，数据集与基准测试赛道成为核心领域，投稿量达 1995 篇，聚焦数据质量提升、基准

测试公平性完善与研究可复现性。同时，AI 安全（鲁棒性、公平性、隐私保护）与 LLM 推理可控性成为刚需，通过测试时优化、门控注意力架构等技术，让模型输出更稳定、可追溯，筑牢学术诚信与技术落地的安全防线。

四、热点论文

本届 NeurIPS 会议共选出 **4 篇最佳论文**（含 1 篇数据集与基准赛道专属获奖论文）。

Artificial Hivemind: The Open-Ended Homogeneity of Language Models (and Beyond)^[1]：大语言模型在生成创意内容时，常因趋向同质化而难以呈现人类风格的多样性，而当前可规模化评估这种多样性的方法仍较为匮乏。为填补这一研究空白，论文作者构建了大规模数据集 Infinity-Chat，该数据集包含 26K 条多样且源自真实世界的开放式用户查询，成为首个支持系统研究真实场景下开放式查询的核心资源。这项研究聚焦现代语言模型的多样性表现、价值多元性及社会影响，通过扎实的数据集构建与系统性分析，为理解和优化语言模型的生成同质化问题提供了关键支撑，具备重要且及时的学术与实践意义。

Gated Attention for Large Language Models: Non-linearity, Sparsity, and Attention-Sink-Free^[2]：随着大语言模型向更大规模、更长上下文方向演进，训练稳定性与注意力行为可控性已成为核心瓶颈。门控机制的有效性虽已得到广泛验证，但在注意力机制中的应用价值及规模化扩展能力尚未被充分探讨。通义千问团队通过一系列控制实验，系统性证实门控机制对大语言模型的优化作用，其核心优势源于增强注意力机制的非线性特征与提供输入相关的稀疏性。此外，该机制还能有效消除注意力池（Attention Sink）和巨量激活（Massive Activation）等问题，显著提升模型训练稳定性，大幅减少训练过程中的损失波动（loss spike），同时在长度外推任务上较基线模型实现明显突破。研究团队在不同尺寸、架构及训练数据规模的模型上验证了方法的通用性，并成功应用于 Qwen3-Next 模型。在当下 LLM 领域科学成果公开共享逐渐减少的环境中，作者公开分享大量工业级计算资源支持下的研究结果，将有力推动社区对大语言模型注意力机制的深入理解，

值得高度称赞。

1000 Layer Networks for Self-Supervised RL: Scaling Depth Can Enable New Goal-Reaching Capabilities^[3]: 该论文聚焦自监督强化学习的可扩展性问题, 深入探索关键积木模块, 指出网络深度是影响性能的核心因素。近年来多数强化学习 (RL) 研究依赖 2-5 层的较浅网络结构, 而本文通过实验证实, 将网络深度提升至 1024 层可显著增强模型性能。研究在无监督的目标条件设定下开展: 不提供示范数据或奖励信号, 要求智能体从零开始探索, 自主学习最大化达成指令目标的概率。在模拟运动控制 (locomotion) 与操作 (manipulation) 任务的评估中, 该方法在自监督对比式 RL 算法上实现 $2\times-50\times$ 的性能提升, 且表现优于其他目标条件基线方法。增加模型深度不仅提高了任务成功率, 还在质量上改变了智能体学到的行为模式。这项研究挑战了“强化学习提供的信息不足以引导深度神经网络中大量参数, 大型 AI 系统应主要依赖自监督训练、RL 仅用于微调”的传统观点, 提出一种新颖且易于实现的 RL 范式, 使极深神经网络的有效训练成为可能, 为强化学习的深度扩展提供了全新思路。

Why Diffusion Models Don't Memorize: The Role of Implicit Dynamical Regularization in Training^[4]: 本文探讨了扩散模型为何能避免对训练数据的死记硬背并具备良好泛化能力, 这一核心问题尚未得到充分解答。该论文聚焦训练动力学在模型从泛化走向记忆转变过程中的作用, 通过大量实验与理论分析, 识别出两个关键时间尺度: 较早的时间点 t_g 与更晚的时间点 t_m 。模型在 t_g 之后开始能够生成高质量样本, 而超过 t_m 后则会出现记忆化 (memorization) 现象。研究围绕扩散模型的隐式正则化动力学展开基础性探索, 将经验观察与形式化理论有机统一, 得出强有力的研究结论, 为理解扩散模型的泛化机制、优化训练过程提供了重要的理论支撑, 对扩散模型的进一步发展具有关键指导意义。

五、大会获奖和竞赛奖

时间检验奖: 时间检验奖是 NeurIPS 的重要荣誉, 旨在表彰 **10 年前发表于 NeurIPS**、经长期实践验证对

人工智能领域产生深远且持续性影响的奠基性研究成果。该奖项无固定首次设立年份, 核心目的是回溯并认可那些突破当时技术局限、塑造后续研究方向, 且在产业落地或学术探索中持续发挥价值的经典工作, 为领域树立“经得住时间考验”的创新标杆。本年度获得该奖的是:

Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks^[5]. 作者: 任少卿、何恺明、Ross Girshick、孙剑。该论文创新性提出“区域提议网络 (RPN)”, 首次将目标检测的“区域提议”与“特征提取、分类回归”整合为端到端深度学习网络, 解决了传统目标检测“速度与精度难以兼顾”的核心痛点——将检测速度从秒级提升至毫秒级, 首次实现实时目标检测, 为计算机视觉技术从实验室走向产业应用奠定基础。

Sejnowski-Hinton 奖: Sejnowski-Hinton 奖首次设立于 2025 年, 其起源于人工智能领域两位先驱——Geoffrey Hinton 与 Terry Sejnowski 的长期合作与学术慷慨。奖项核心目的是表彰**过去 10 年内发表**、在“基于人工智能洞见的大脑计算理论”领域做出重大贡献, 且对 NeurIPS 社区产生显著影响的研究, 以此纪念二人的合作精神, 并推动“脑启发人工智能”与“计算神经科学”的交叉发展。本年度获得该奖的是:

Random Synaptic Feedback Weights Support Error Backpropagation for Deep Learning^[6]. 作者是 Timothy Lillicrap、Daniel Counden、Douglas Tweed、Colin Akerman。论文聚焦“大脑神经回路如何仅通过局部信息高效学习”这一经典问题, 提出“反馈对齐 (Feedback Alignment)”机制——证明多层人工神经网络无需像反向传播那样依赖“精确的反向权重对称性”, 仅通过固定的随机反馈权重即可有效学习。训练过程中, 网络的前向权重会自然与随机反馈信号对齐, 生成有用的损失梯度估计, 首次为“权重传输问题”(真实神经元如何在无跨区域信息传递的情况下遵循损失梯度) 提供了符合生物学合理性的具体解决方案。

六、总结展望

NeurIPS 2025 继续扮演全球 AI 研究的“风向标”。从本届日程与教程主题可见，研究重心正从单纯的规模扩张转向“在效率与可靠性约束下提升智能”：算力与能耗被当作一等设计指标，对齐与安全、可解释/可审计、隐私与合规等问题被更前置地讨论。

展望未来，架构创新将更紧密地与系统工程、数据与评测方法协同推进：一方面缓解长上下文、训练/推理成本与真实场景泛化的瓶颈，另一方面把可控性、可追溯性与人类价值约束内生到模型与流水线中，推动 AI 从“能跑分”走向“更稳、更省、更可用”规模化落地。

责任编辑 魏秀参

参考文献

- [1] Liwei Jiang, et al. Artificial Hivemind: The Open-Ended Homogeneity of Language Models (and Beyond). NeurIPS 2025.
- [2] Zihan Qiu, et al. Gated Attention for Large Language Models: Non-linearity, Sparsity, and Attention-Sink-Free. NeurIPS 2025.
- [3] Kevin Wang, et al. 1000 Layer Networks for Self-Supervised RL: Scaling Depth Can Enable New Goal-Reaching Capabilities. NeurIPS 2025.
- [4] Tony Bonnaire, et al. Why Diffusion Models Don't Memorize: The Role of Implicit Dynamical Regularization in Training. NeurIPS 2025.
- [5] Shaoqing Ren, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. NeurIPS 2015.
- [6] Timothy P. Lillicrap, et al. Random Synaptic Feedback Weights Support Error Backpropagation for Deep Learning. Nature Communications 2016.



叶翰嘉

现任南京大学人工智能学院副教授、博导，在南京大学机器学习与数据挖掘研究所（LAMDA）从事学术研究工作；主持国家自然科学基金青年科学基金 B 类项目、国家重点研发计划专项项目，获中国人工智能学会吴文俊人工智能青年科技奖。

Email: yehj@nju.edu.cn