顶会观察

# NeurIPS 2023

上海交通大学　陈思衡　雷梓行　魏思哲

神经信息处理系统大会（Conference on Neural Information Processing Systems, NeurIPS）是机器学习领域的顶级会议，是神经计算方面最好的会议之一，在中国计算机学会推荐国际学术会议中被评为人工智能领域的 A 类会议。在 Google Scholar 发布的 2020 年学术指标中，h5 指数高达 309，位于计算机领域的第 2 位、位列所有领域出版物的第 9 位。

今年第 37 届 NeurIPS 于 12 月 10 日到 16 日在美国路易斯安娜州的新奥尔良举办。

## 一、NeurIPS 2023 的亮点

今年的 NeurIPS 是新冠疫情流行结束后第一届 NeurIPS 大会，参会人数大量增加，大会在巨大的新奥尔良 Ernest N. Morial 会议中心举办，有着巨大的 Poster 展区和多个报告厅，能够容纳数千名参加者。大会采用 Whova App 进行组织，参会者能够从手机实时查询各个报告、研讨会、海报的所在位置和参与嘉宾，同时可以非常轻松地将相关日程加入日历提醒，方便安排会议行程。同时，大会主办方也鼓励大家通过 App 进行社交活动，自由地创建群组，并寻找潜在的合作者。

除了论文成果展示之外，NeurIPS 的一大特色是精心组织的社交活动。这些社交活动有着丰富多彩的目标人群，例如女性研究者、交叉背景的研究者、医疗健康从业人群等等。这些社交活动提供了一些非正式的场合让各种身份的研究者们互相了解，分享经验，增进友谊。

## 二、论文录用情况

NeurIPS 2023 共有 13330 篇论文被提交，创历史新高。最终有 3540 篇论文入选，录取率不到 26.6%，略高于去年的 26.1%。其中 77 篇被录用为 Oral，400 篇被录用为 Spotlight，Oral 和 Spotlight 的录取率仅有 0.578%和 3.00%。

在 NeurIPS 2023 ，据不完全统计，谷歌位于所有高校/机构的榜首，入选论文高达 180 篇。斯坦福大学与麻省理工学院并列排名第二，共有 130 篇论文入选。卡耐基梅隆大学位居第三，共有 112 篇论文入选。国内高校/机构中，清华大学排名第一，共有 111 篇论文入选，排在国内外所有高校/机构的第 4 位；北京大学第二，共有 98 篇论文入选，排在国内外所有高校/机构的第 5 位。入选论文较多的国内机构还有中国科学院、上海交通大学、香港中文大学、浙江大学和中国科学技术大学。

NeurIPS 作为机器学习领域的顶级会议，收录的论文包罗了人工智能领域的各种主题，包括大语言模型，人工智能生成内容（AIGC），深度学习及其应用，强化学习和规划，计算机视觉，纯理论研究、概率方法、优化，机器学习和社会，神经科学和认知科学等方方面面。特别是在今年生成式人工智能的浪潮下，大语言模型以及相关工作受到了广泛的关注。

## 三、邀请报告

NeurIPS 2023 共有七个邀请报告（Invited Talk），主题丰富，涵盖了火热的生成式人工智能，负责任的人工智能，神经认知科学，机器学习系统，大语言模型和强化学习与数字医疗等，反映了机器学习正与其他学科不断相互影响并相互交融；同时探讨了人工智能大潮下的多个社会问题，展现了科研人员的社会责任感。慕尼黑大学的正教授 Björn Ommer 介绍了他对规模扩张的

错觉与生成式人工智能的未来的看法。Google Research 的研究科学家 Lora Aroyo 呼吁机器学习系统应当考虑内容的固有模糊性和人类观点的自然多样性，应当建立对文化意识和以社会为中心的研究，关注数据质量和数据多样性的影响，用于训练和评估机器学习模型，并在不同的社会文化环境中促进负责任的人工智能部署。印第安纳大学布卢明顿分校的杰出教授，认知科学和认知发展领域国际公认的领导者 Linda Smith 报告了她采用复杂系统的视角，旨在理解感知、运动和认知发展在产后前三年间的相互依赖性的研究。加州大学伯克利分校电气工程与计算机科学系的教授 Jelani Nelson 介绍了数据草图这一概念。数据草图对内存进行压缩的总结，但仍然允许回答有用的查询；作为一种工具，在算法设计、优化、机器学习等领域都有所应用。斯坦福大学人工智能实验室的副教授 Christopher Ré 介绍了他关于基础模型的研究，着重描述了一种基于经典信号处理的新型架构。哈佛大学统计与计算机系的教授 Susan Murphy 介绍了她的小组在开发在线强化学习（RL）算法以应用于数字健康干预中所面临的一些挑战及初步解决方案，用于帮助那些正在努力应对诸如物质滥用、高血压和骨髓移植等健康问题的患者。在“Beyond Scaling Panel”，来自 Google Brain 的 Aakanksha Chowdhery、康奈尔大学的 Alexander Rush、Meta AI 的 Angela Fan、清华大学的唐杰教授以及斯坦福大学的 Percy Liang 分享了他们对大模型的看法和思考。

## 四、会议热点论文

本次会议涌现了许多优秀的工作，它们具有非常高的学术价值与应用价值。NeurIPS 2023 共有 6 篇获奖论文，其中包含杰出论文（2 篇）、杰出论文亚军（2 篇）和杰出数据集和基准论文（2 篇）。值得一提的是，六篇获奖论文中有四篇是关于大语言模型，这也凸显了研究人员们对这一领域的重视。荣获杰出论文大奖的两篇论文分别是自 Google Deepmind 团队的 Privacy Auditing with One (1) Training Run、斯坦福的 Are Emergent Abilities of Large Language Models a Mirage?。

**Privacy Auditing with One (1) Training Run.** 这项工作提出了一种只需要单次训练来检查隐私机器学习系统的方案。相比现有方法动辄需要数百个训练模型，论文中提出的方案仅需单次训练，其基于差分隐私机器学习系统能够独立添加或删除多个训练示例的并行性的特点，分析了差分隐私和统计泛化的联系，避免了群体隐私的成本。该方法对算法的假设要求很低，可以应用于黑盒或者白盒环境。

**Are Emergent Abilities of Large Language Models a Mirage?** 斯坦福的研究人员们在这篇工作中针对大语言模型的“涌现”能力提出了一种新的解释。针对“涌现”能力的突现性和不可预测性，该文认为在特定任务和模型中研究者选择了特定的度量标准。具体来说，非线性或者不连续的度量会产生明显的“涌现”能力，而线性或者连续度量则会产生平滑、连续、可预测的模型性能变化。该文通过三种互补的方式针对包括 GPT-3 系列中声称具有“涌现”能力的任务进行了检验，以及展示了如何选择度量标准从而在视觉任务中创造“涌现”能力。该文证明“涌现”能力与度量或者统计标准有关，而不是人工智能的基本属性得到了扩展。

杰出数据集论文奖由 ClimSim: A large Multi-scale Dataset for Hybrid Physics-ML Climate Emulation 获得。该数据集由来自于加州大学尔湾分校、哥伦比亚大学、英伟达等 20 个机构的气候科学家和机器学习研究人员共同发布。文章中发布的数据集 ClimSim 是一个用于混合物理-ML 气候仿真的大型多尺度数据集，将物理学与机器学习相结合引入了新一代更高保真度的气候模拟器，通过机器学习模拟器执行计算密集、短时、高分辨率的任务，从而绕开摩尔定律。借助该数据集，有望提高诸如风暴等气象预测的准度与精度，造福人类社会。

杰出基准论文奖颁给了 DECODINGTRUST: A Comprehensive Assessment of Trustworthiness in GPT Models。该基准由伊利诺伊大学香槟分校、斯坦福大学、加州大学伯克利分校等机构共同发布，其提出了一套针对大语言模型的全面可信度评估框架，包括毒性、刻板印象、对抗性攻击鲁棒性、分布外鲁棒性、对抗性示范鲁棒性、隐私、机器伦理和公平性等维度，并

发现了之前未被公开的有关信任度的漏洞。该基准框架的发布有助于 GPT 模型在更多领域的应用。

杰出论文亚军分别是来自 Hugging Face、哈佛大学、图尔库大学的 Scaling Data-Constrained Language Models 和来自斯坦福大学与 CZ Biohub 的 Direct Preference Optimization: Your Language Model is Secretly a Reward Model。前者探讨了在数据有限的情况下大语言模型的扩展，后者介绍了一种改进大语言模型行为以符合人类偏好的新方法——直接偏好优化。

除了六篇获奖论文之外，今年的时间检验奖颁给了 Distributed Representations of Words and Phrases and their Compositionality。该论文由十年前还在 Google 从事研究的 Tomas Mikolov，Ilya Sutskever，Kai Chen，Greg Corrad，Jeffrey Dean 等人撰写，被引量已经超过了 4 万次。论文作者团队中 Greg 与 Jeffrey 也来到现场做了演讲，并分享了这篇工作诞生的经历。官方给出的颁奖理由是：这项工作引入了开创性的词嵌入技术 word2vec，展示了从大量非结构化文本中学习的能力，推动了自然语言处理新时代的到来。可谓实至名归。

NeurIPS 2023 一共收录了 77 篇 Oral 论文和 400 篇 Spotlight 论文，大会将 Oral 论文分成了 22 场 talk，包括：RL, Datasets & Benchmarks, Tractable models, DL Theory, Efficient Learning, Objects/Neuroscience/Vision, Causality, Privacy, Neuro, NLP / Tools, Diffusion Models, Optimization, COT / reasoning, GNNs / Invariance, Privacy / Fairness, Probability / Sampling, Vision, LLMs, Theory 等方面。

Oral 论文仅仅是录用论文的冰山一角。其他录用论文中还有大量值得探索和讨论的内容，毕竟 word2vec 并不是那一年的 Oral 或者 Spotlight，甚至其原始论文还曾遭遇 ICLR 拒稿，可见非 oral 的录用论文中也是藏龙卧虎。

海报展览在应用、深度学习、机器学习、优化、概率方法、强化学习、社会影响、理论等八大领域中展开，分为六个时间段进行。每个时间段仅有两小时展示时间，让我们难以全面、细致地了解每一篇研究论文。虽然 NeurIPS 偏重理论研究，但人工智能应用领域的研究同样激发了研究人员的浓厚兴趣。像 ChatGPT 这样的成功案例更是吸引了众多研究者，促使他们思考如何将尖端技术应用于实际。本次大会中两个尤为引人瞩目的方向是 AI Safety 和 AI Agent。在主旨演讲以及多个 workshop 上呼吁更多的研究者能够关注到大语言模型时代愈发突出的 AI 安全性问题，主题包括如何避免偏见、与人类对齐、数据安全和攻击防护等诸多方向。AI agent 则是另外一个引人注目的方向，年初斯坦福大学的 AI 小镇风靡全球，也吸引了更多的研究者关注到这个方兴未艾的领域。在 Foundation Models for Decision Making Workshop 中，来自全球各地的研究者分享了他们在 AI agent 方面的令人激动的研究成果。例如，Percy Liang 教授从 AI 小镇开始，描述了在多智能体社会中信息“Diffusion”，给我们多智能体交互的研究甚至是社会学研究提供了新的思路和方向。Liang 教授研究组关于 research AI 的研究也非常有趣，在大语言模型时代，我们也许应该将我们的思维从“如何做一件事”转向“如何让 AI agent 学会帮我做这件事”。CMU 的 Ruslan Salakhutdinov 教授介绍了他的研究组关于 Web AI agent 的研究，通过增加模型多模态能力和思维链能力，AI agent 可以不断进化，帮助人类完成许多日常的任务，例如订机票，交电费，搜索汇总资料等等。AI agent 很有可能在未来几年内深刻影响人类的数字生活，为每个人都配备善解人意、随叫随到的“贾维斯”。学科交叉也是 NeurIPS 的鲜明特征，这次的 NeurIPS 上有许许多多的交叉学科研究。有很多的研究者创造性地将大模型应用在各个领域，包括用于自动驾驶的 DriveGPT，用于地理领域的 GPT4GEO 等等。神经科学与 AI 的交叉也受到了广泛的关注，大会中有数篇论文关注到了通过 Diffusion Model 将 MRI 信号解码成视觉图像，帮助我们更深刻地理解人类大脑。如同 AlphaFold 深刻地影响了结构生物学，随着 AIGC 日新月异的发展，越来越多的领域有可能借助 AI 的力量，找到新的方向和机遇。

除此之外，大会还举办了 14 个教程（Tutorials）和 58 个研讨会（Workshops），包括语言模型、世界模型、Diffusion Model 等丰富的领域。这些教程和研讨会的安排非常紧凑，在同一时间段内甚至有多场供参会人员选择，为科研人员扩展视野和交流互动提供了便利的渠道和平台，利于机器学习领域的长期发展。

## 五、总结与展望

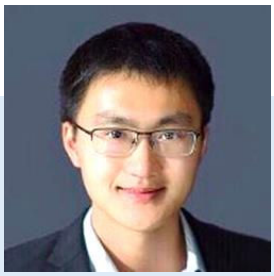
在 NeurIPS 2023 录用的 3584 篇论文关键词中，“Bayes”（174 次）、“Gaussian”（195 次）出现的频次仍然很高，说明经典的方法与理论研究仍然在该会议中受到青睐。与此同时，一些时下热门研究方向在 NeurIPS 上也能够大放异彩。据不完全统计，“Generative”、“Transformer”、“Agent”、“Zero/Few-shot”等词语分别出现在 508、271、280、211 篇接收论文关键词中，在会场中也能感受到 LLMs、Diffusion Models 等领域受到了广泛的关注。此外，尽管是一场机器学习领域的会议，视觉领域和机器人领域的工作也非常多，在教程和研讨会上有相当数量的参与者，共同讨论着各个应用领域的热门话题。

参与疫情后首个全面面对面的会议，与来自世界各地的研究者进行直接的沟通与交流，这种体验远胜于线上虚拟会议。面对面的对话和休闲闲谈在会场中激发了更多的科研创意，也促成了与全球顶尖科研人才的合作。因此，尽管线上会议有其成本效益，但面对面会议带来的益处是不言而喻的。

责任编委 王金甲

## 参考文献

[1] https://neurips.cc/virtual/2023/papers.html?mode=detail&filter=titles
[2] https://blog.neurips.cc/
[3] https://voxel51.com/blog/neurips-2023-and-the-state-of-ai-research/



**陈思衡**

上海交通大学未来媒体网络协同创新中心长聘副教授，国家高层次人才青年项目。研究方向为多智能体协作学习。团队信息见网站 https://siheng-chen.github.io。
Email: sihengc@sjtu.edu.cn

**雷梓行**

上海交通大学硕士生，师从陈思衡教授。研究兴趣包括智能体交互，具身智能，更多信息详见：https://chezacar.github.io/。
Email: chezacarss@sjtu.edu.cn

**魏思哲**

上海交通大学硕士生，师从陈思衡教授、张娅教授。研究方向为自动驾驶、协作感知。个人网站：https://sizhewei.github.io。
Email: sizhewei@sjtu.edu.cn