



大模型时代的NLP博士之路： 跨学科视角下的成长与转型经 验谈

聂耳聪

Center for Information and Language Processing,
Ludwig Maximilians University of Munich (LMU)
<https://ercong21.github.io/>

June 22, 2025

About me

聂耳聪 PhD Candidate

Schütze NLP Lab, Center for Information and Language Processing (CIS),
Ludwig Maximilians University of Munich (LMU Munich),
Munich Center for Machine Learning (MCML)

- 慕尼黑大学(LMU Munich)信息与语言处理中心(CIS)自然语言处理(NLP)方向的博士毕业年级学生,
- 隶属于前ACL主席Hinrich Schütze教授领导的Schütze NLP Lab, 同时为慕尼黑机器学习中心(MCML)青年成员,
- 硕士(M.Sc.): 计算语言学与计算机科学硕士学位, 慕尼黑大学,
- 本科(B.A.): 德语与金融学士学位, 上海交通大学,
- 研究兴趣包括: 多语言自然语言处理、高效NLP方法以及受人类启发的自然语言处理



<https://ercong21.github.io/>

About Schütze NLP Lab

Welcome to the Schütze Lab

We are a dynamic research group at the [Center for Information and Language Processing](#) at [Ludwig Maximilian University Munich](#), under the supervision of [Prof. Hinrich Schütze](#). Our research areas include:

- **Large Language Models (LLMs)**: We explore the behavior, structure, and potential of LLMs, examining their capabilities, biases, and self-assessment mechanisms to improve reliability and interpretability.
- **Knowledge Expansion in NLP Models**: We investigate how models can acquire and integrate new knowledge over time, using techniques that help improve their comprehension and generation abilities.
- **Representation Learning and Interpretability**: We study how language models represent linguistic and conceptual information by analyzing neurons and internal circuits to better understand and refine model behavior.
- **Multilingual NLP**: We address challenges in processing and evaluating multiple languages by developing benchmarks and methods for multilingual evaluation, including work on low-resource languages.
- **Intersection of NLP and Robotics**: We integrate language understanding into robotic systems to enable natural, adaptable interaction in multimodal environments.

You can find resources (data, code, repositories) released by CIS lab members on our [GitHub page](#) and [HuggingFace page](#).



Lab page:
<https://cislsp.github.io/>

About MCML

慕尼黑机器学习中心

Munich Center for Machine Learning,
MCML

<https://mcml.ai/>

MCML

A top spot for AI and ML research in Europe

We attract the most promising researchers from all career levels to our center and transfer innovations and AI potential to industry and society.

Research Activities

The research of MCML is concentrated in three main areas.

A | Foundations of ML serves as the groundwork for methodological advancements in ML.

B | Perception, Vision, and NLP forms a dynamic research domain at the intersection of computer science and cognitive sciences.

C | Domain-Specific ML facilitates the development of ML methods to address application-specific and socially significant issues, working closely with the other two areas.

目录

一、个人研究背景和兴趣

二、NLP博士生涯的大模型冲击

三、跨学科与交叉创新重要性

四、交流讨论

目录

一、个人研究背景和兴趣

二、NLP博士生涯的大模型冲击

三、跨学科与交叉创新重要性

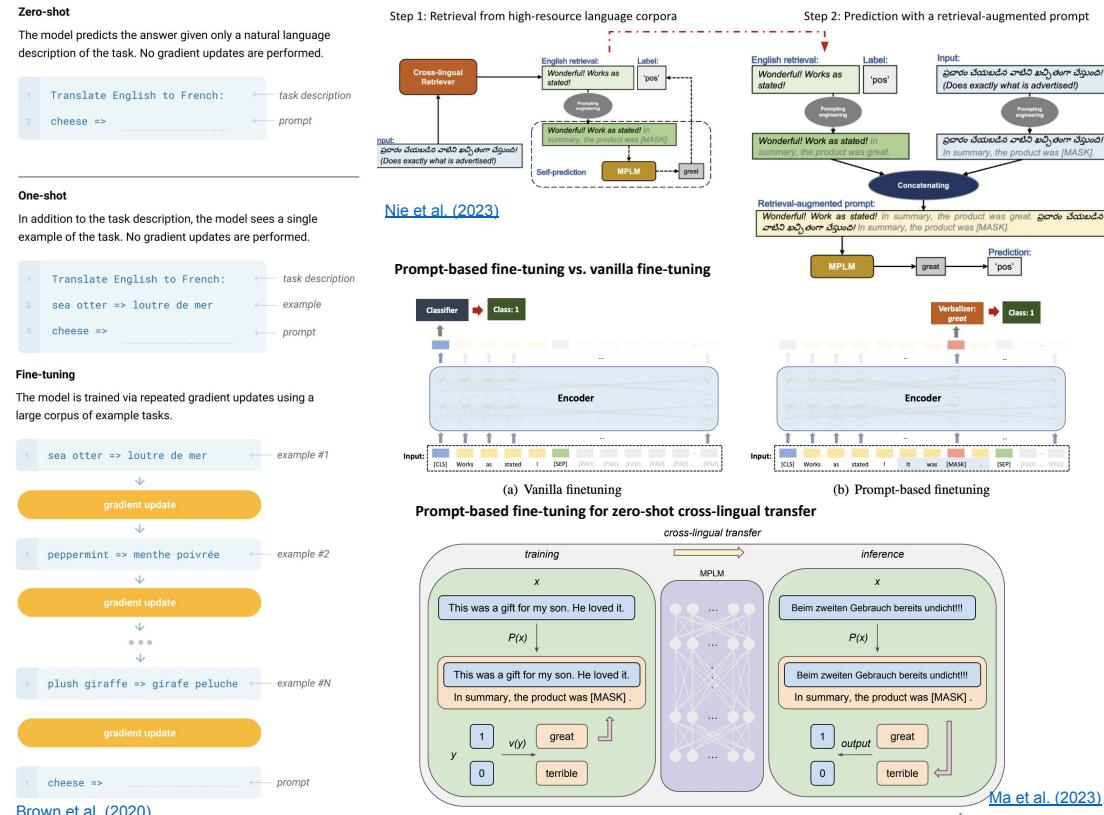
四、交流讨论

Research Profile(1)

多语言NLP与跨语言迁移: 致力于探索基于提示的学习(prompt-based learning)等新范式在多语言和低资源语言场景下的迁移能力, 提出了跨语言检索增强提示(PARC)等方法, 提升了大模型在低资源语言上的零样本表现, 并系统研究了大模型的多语言能力与迁移机制。

高效NLP方法:关注低资源数据条件下的数据增强、参数高效微调(PEFT)、模型剪枝等方向,聚焦NLP模型在资源受限场景下的实用性和可扩展性。

人类启发与可解释NLP:结合计算神经语言学与人类语言处理机制,研究大模型内部的语言结构与知识表征,致力于大模型的可解释性和跨学科应用。



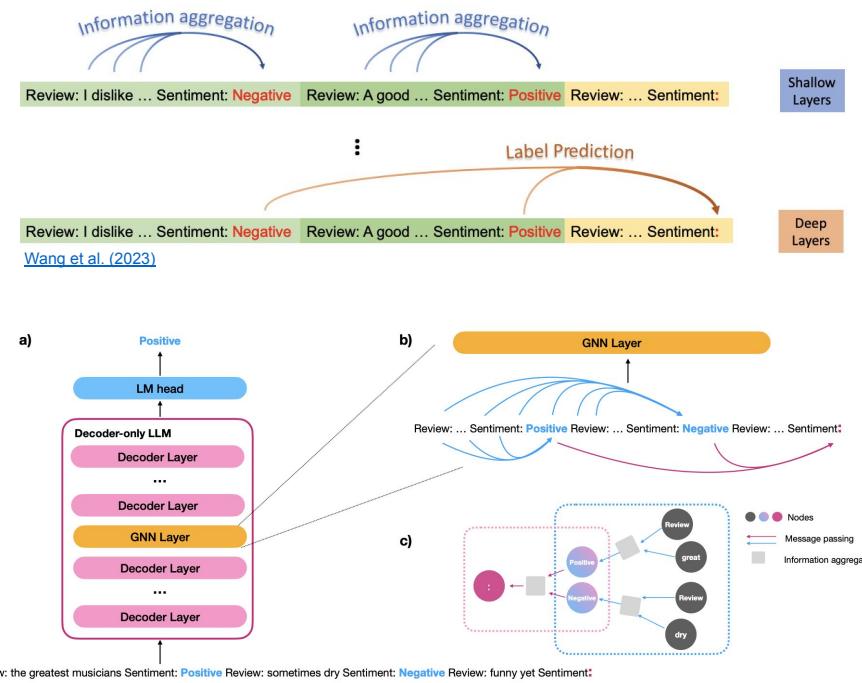
Research Profile(2)

多语言NLP与跨语言迁移: 致力于探索基于提示的学习(prompt-based learning)等新范式在多语言和低资源语言场景下的迁移能力, 提出了跨语言检索增强提示(PARC)等方法, 提升了大模型在低资源语言上的零样本表现, 并系统研究了大模型的多语言能力与迁移机制。

高效NLP方法: 关注低资源数据条件下的数据增强、参数高效微调(PEFT)、模型剪枝等方向, 聚焦NLP模型在资源受限场景下的实用性和可扩展性。

人类启发与可解释NLP: 结合计算神经语言学与人类语言处理机制, 研究大模型内部的语言结构与知识表征, 致力于大模型的可解释性和跨学科应用。

GNNavi:一种受上下文学习信息流启发的图神经网络参数高效微调方法



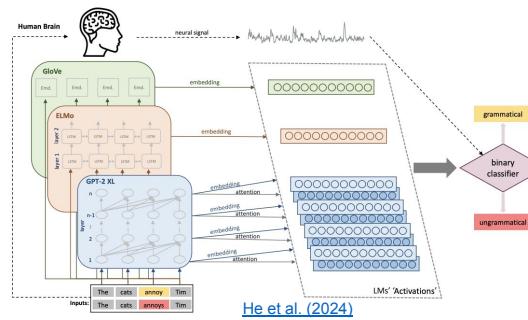
Research Profile (3)

多语言NLP与跨语言迁移: 致力于探索基于提示的学习(prompt-based learning)等新范式在多语言和低资源语言场景下的迁移能力, 提出了跨语言检索增强提示(PARC)等方法, 提升了大模型在低资源语言上的零样本表现, 并系统研究了大模型的多语言能力与迁移机制。

高效NLP方法: 关注低资源数据条件下的数据增强、参数高效微调(PEFT)、模型剪枝等方向, 聚焦NLP模型在资源受限场景下的实用性和可扩展性。

人类启发与可解释NLP: 结合计算神经语言学与人类语言处理机制, 研究大模型内部的语言结构与知识表征, 致力于大模型的可解释性和跨学科应用。

Probing



Neuron Attribution and Intervention

User Prompt



ما الفرق بين السور المكية والمدنية في القرآن الكريم؟
English translation: What is the difference between the Meccan and Medinan surahs in the Holy Quran?

LLM Response

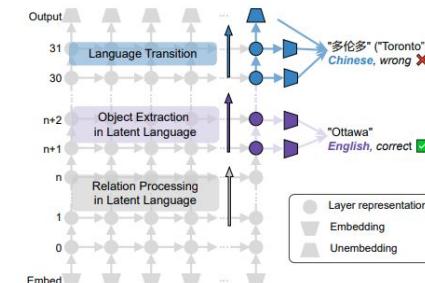


The difference between Meccan and Medinan surahs in the Quran is a significant aspect of Islamic studies. Here's a brief overview:
Meccan Surahs:
1. **Time period:** The Meccan surahs were revealed during the early years of Prophet Muhammad's prophethood, ...

Mechanistic understanding and mitigation of language confusion in LLMs (Nie et al., 2025)

LogitLens

Output	文	:	-"	花
31	文	:	-"	花
29	文	:	-"	花
27	文	:	_flower	花
25	文	:	_flowe...	_flowe...
23	文	:	-"	_flowe...
21	文	:	_flowe...	_flowe...
19	文	:	-"	_flowe...
17	eval	:	-"	<0xE5>
15	ji	:	-"	Ψ
13	i	_vac	ois	_bore
11	eda	eda	_Als	abel
9	eda	ná	_Als	_hel
7	iser	arie	◀	arias
5	npa	orr	◀	arias
3	心	ures	_Bedeut	arda
1	_beskre	化	Portail	_Kontr...
	中	文	:	-"



加拿大的首都在哪里? 答案是:
(What is the capital of Canada? The answer is:)

Cross-lingual factual recall inconsistency (Wang et al., 2025)

目录

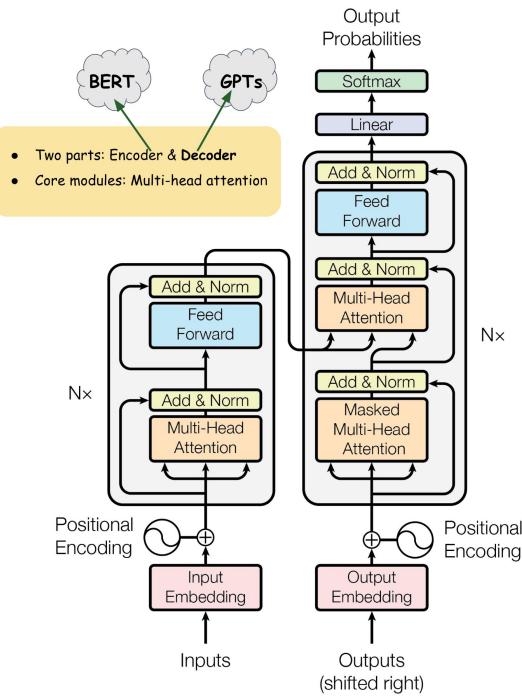
一、个人研究背景和兴趣

二、NLP博士生涯的大模型冲击

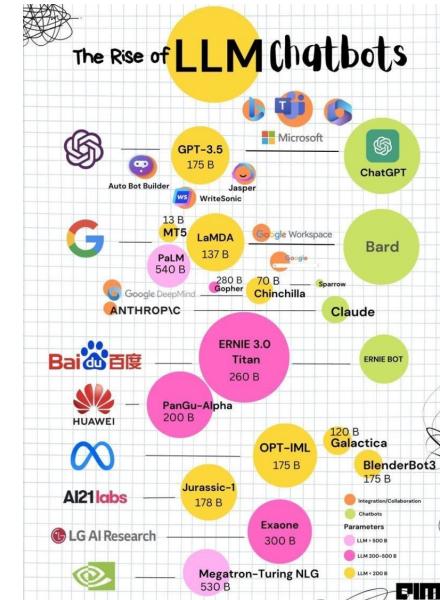
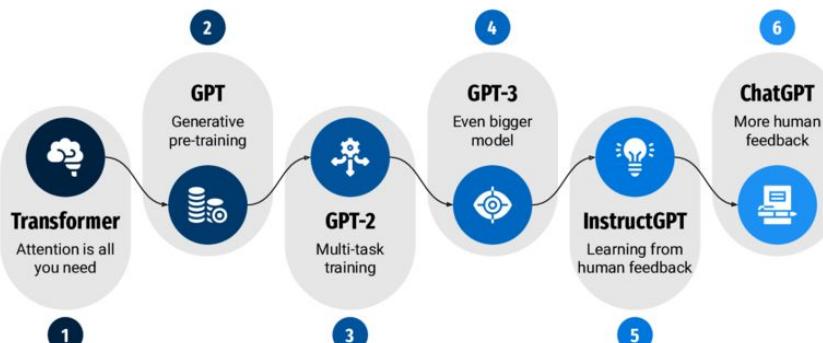
三、跨学科与交叉创新重要性

四、交流讨论

大语言模型的出现



Evolution from Transformer architecture to ChatGPT



大语言模型对NLP博士研究的冲击

博士开始时的NLP研究范式

- 2022年博士入学时, NLP主流仍以BERT等Encoder架构为核心
- GPT-3虽已问世, 但未开源, 影响有限。
- Prompt-based learning刚刚兴起, 主要作为新颖的研究范式被关注。
- 博士初期研究聚焦基于提示学习方法的多语言NLP、跨语言迁移, 均基于Encoder模型(如mBERT, XLM-R)。

大模型(如ChatGPT)出现后的冲击与转折

- 2022年底ChatGPT横空出世, LLM展现出强大的上下文学习、推理与生成能力, 随着10B级别左右的开源模型越来越多, NLP研究生态彻底改变。
- Prompt-based learning成为主流范式, 围绕LLM强大的生成、泛化、知识、推理等能力成为研究新热点。

大模型带来了研究范式的巨大转变:

- 新技术/新范式: RLHF、对齐、强化学习、推理能力、CoT等。
- 新应用: LLM-as-a-judge、LLM-as-annotator、合成数据生成、基于prompt的probing等。
- 新挑战: 大模型的可解释性、评测基准集升级、高效微调(PEFT)、KV-Cache等。

个人研究方向的调整与思考

个人研究方向的调整经历

- 初期对大模型持怀疑和观望态度，继续深耕传统多语言NLP(又做了一年“传统”方向)。
- 2024年起逐步拥抱大模型，研究重心转向LLM的多语言能力、可解释性与跨学科结合。

人类启发的NLP研究(Human-inspired NLP)

- 另一方面，大模型的出现使得人工智能与人类智能之间的关系变得更加微妙和紧密。
- 大模型在外部行为(如推理、语言习得等)和内部机制(如信息表征、神经元激活)上，展现出与人类认知过程的诸多相似性。
- 这为NLP研究带来了全新的视角：不仅关注模型的工程性能，更关注其“认知合理性”和“可解释性”。

主要研究内容与方向：人类启发的大模型的多语言能力与可解释性机制分析

- 探索LLM在多语言环境下的知识表征、迁移与推理机制，分析其在不同语言间的泛化能力和一致性。
- 研究大模型内部如何编码和处理不同语言的结构与语义，揭示其“语言共性”与“语言特性”。
- 关注大模型在处理复杂推理、语言习得等任务时，是否展现出类似人类的认知路径和偏好。

个人研究方向的调整与思考

反思发现：

- 许多“传统”研究方法(如提示校准、跨语言检索增强)在大模型时代依然适用，甚至更具价值(PARC)。
- 大模型的本质仍是Transformer架构，许多底层机制和方法具有延续性，如涉及到模型内部参数和结构的研究方向，包括可解释性(probing)、参数高效微调、注意力机制KV缓存等。
- 抓住大模型带来的“变”与“不变”，在原有多语言NLP基础上，结合大模型新能力，探索跨语言迁移、可解释性等新问题。
- 关注大模型与人类智能的关系，结合计算神经语言学、人类认知等交叉领域，聚焦人类启发的NLP研究。

经验与思考

- 面对范式转型，既要敏锐捕捉新趋势，也要善于挖掘自身领域的**独特优势**。
- 不必盲目追逐热点，应结合个人背景和兴趣，找到与大模型结合的**差异化创新点**，形成独特竞争力。
- **跨学科视角和交叉创新**是大模型时代NLP研究的重要突破口。

目录

一、个人研究背景和兴趣

二、NLP博士生涯的大模型冲击

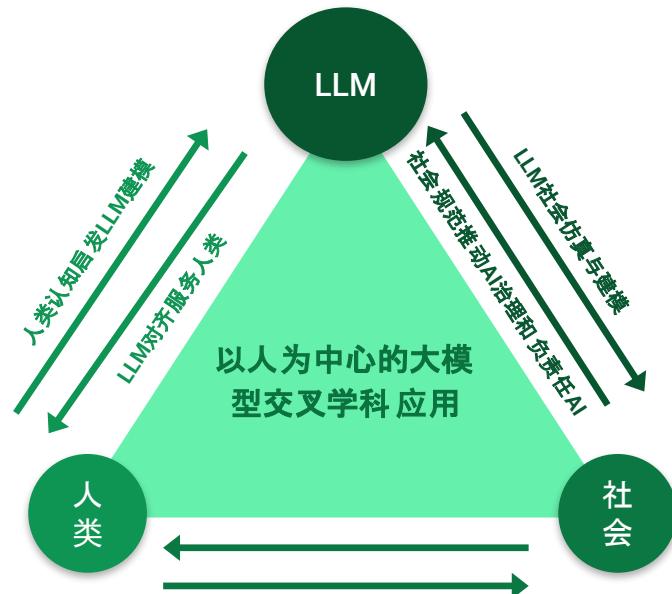
三、跨学科与交叉创新重要性

四、交流讨论

以人为中心的LLM+

大模型与各领域结合的潜力

- LLM具备强大的知识整合、推理和生成能力，为多学科研究带来新范式。
- 跨学科合作可拓展研究视野，提升创新深度和实际影响力。



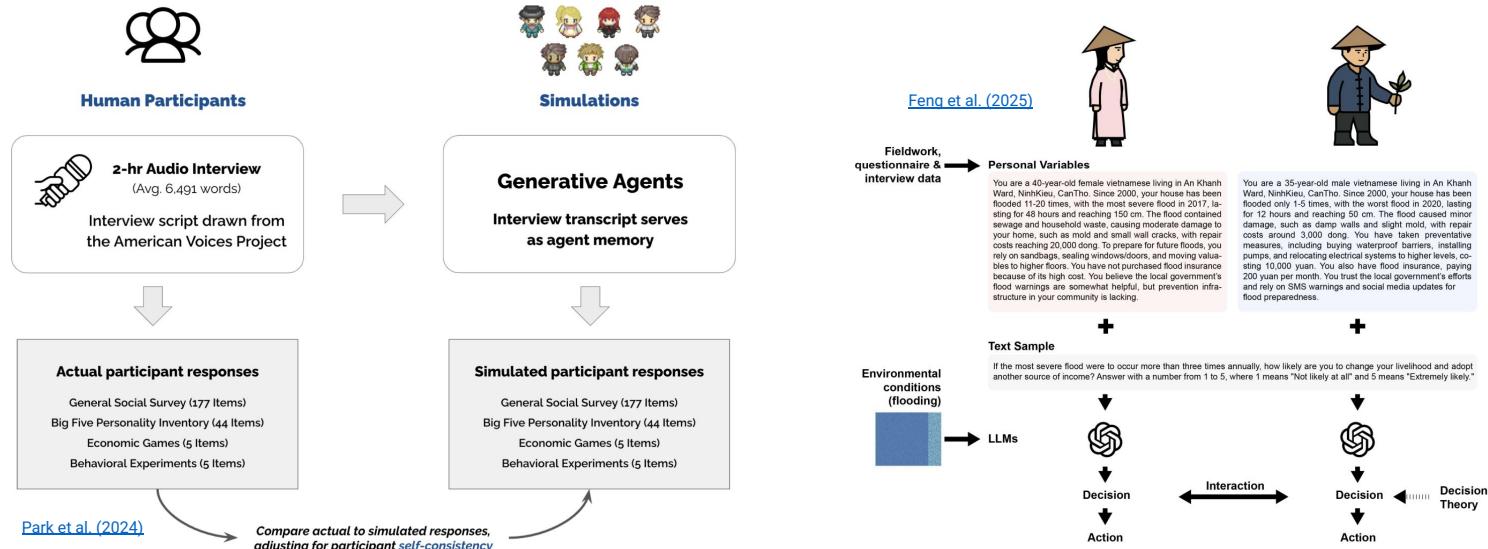
LLM persona与社会科学

利用LLM persona模拟, 进行社会调查、问卷研究、社会建模。

人格化、文化敏感的
LLM persona

- 机器心理学
(Machine Psychology)

[Hargendorff et al. \(2024\)](#)

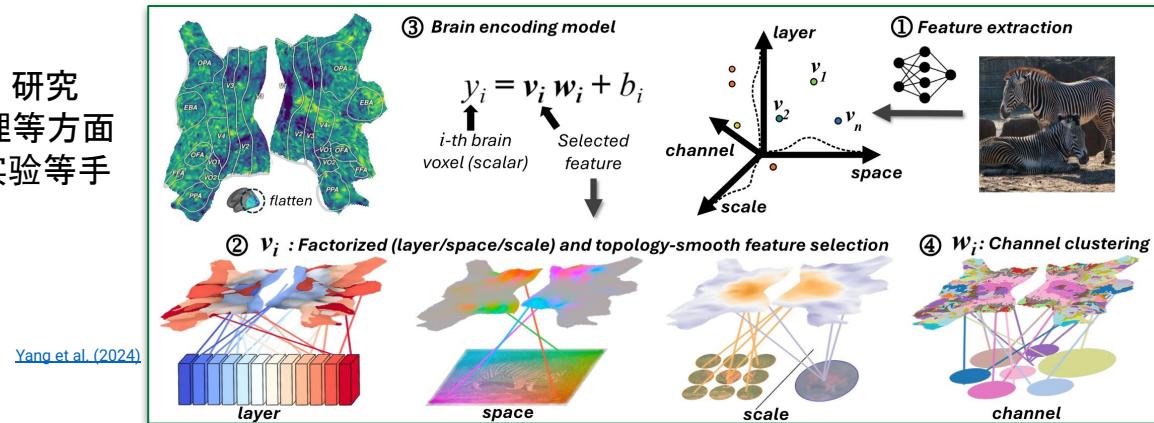


典型案例: 斯坦福“Generative Agent Simulations”——1000+ LLM personas模拟虚拟社会行为。

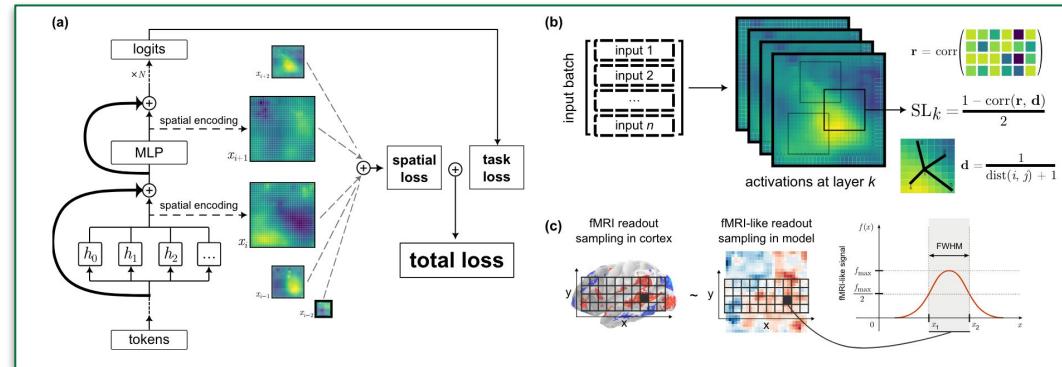
社会建模:结合地理、气候等领域, 模拟灾害情境下的人类反应与社会动态。

LLM与神经科学/脑科学

- 结合脑科学、神经科学方法，研究LLM与人脑在语言处理、推理等方面的不同，通过脑成像、认知实验等手段，探索LLM的“类脑”机制。



- 语言模型的类脑建模：
大脑结构启发语言模型的建模



跨学科视角下重塑NLP的语境

- LLM for mental health,
用于语言障碍检测(如
阿尔茨海默症等)、认知
障碍机制分析
- Clinical NLP, Medical NLP
- Legal NLP
- LLM for Education
- LLM for humanities
- ...

EMNLP 2025 Theme Track: **Interdisciplinary Recontextualization of NLP**

The core interests of the ACL community are rooted in human-language technologies but also have broad reach into other fields. A couple of recent examples are the burgeoning areas of Code models and Vision models. Earlier cases are exemplified through SIGs connected with the fields of education, medicine, and humanities. Movements such as NLP for Social Good and Computational Social Science show a desire for broad impact, which requires expertise beyond the borders of our own community to achieve. This year's theme of **Advancing our Reach: Interdisciplinary Recontextualization of NLP** aims to highlight this need for broader connections with other fields to understand and intensify NLP's impact. The goal is to increase our awareness of how advances in NLP can impact other fields, and design better strategies to measure that impact both within and across disciplines.

References

- Brown, Tom, et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.
- Nie, Ercong, et al. "Cross-Lingual Retrieval Augmented Prompt for Low-Resource Languages." *Findings of the Association for Computational Linguistics: ACL 2023*. 2023.
- Ma, Bolei, et al. "Is Prompt-Based Finetuning Always Better than Vanilla Finetuning? Insights from Cross-Lingual Language Understanding." *KONVENS*. 2023.
- Wang, Lean, et al. "Label Words are Anchors: An Information Flow Perspective for Understanding In-Context Learning." *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. 2023.
- Yuan, Shuzhou, et al. "GNNavi: Navigating the Information Flow in Large Language Models by Graph Neural Network." *Findings of the Association for Computational Linguistics ACL 2024*. 2024.
- He, Linyang, et al. "Decoding Probing: Revealing Internal Linguistic Structures in Neural Language Models Using Minimal Pairs." *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. 2024.
- Nie, Ercong, Helmut Schmid, and Hinrich Schütze. "Mechanistic Understanding and Mitigation of Language Confusion in English-Centric Large Language Models." *arXiv preprint arXiv:2505.16538* (2025).
- Wendler, Chris, et al. "Do llamas work in english? on the latent language of multilingual transformers." *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2024.
- Wang, Mingyang, et al. "Lost in multilinguality: Dissecting cross-lingual factual inconsistency in transformer language models." *Proceedings of ACL 2025*. 2025.
- Park, Joon Sung, et al. "Generative agent simulations of 1,000 people." *arXiv preprint arXiv:2411.10109* (2024).
- Feng, Wenhan, et al. "Generative agent-based modeling for climate adaptation policy: A flood resilience perspective." *Proceedings of Social Simulation Conference 2025*. 2025.
- Hagendorff, Thilo. "Machine psychology: Investigating emergent capabilities and behavior in large language models using psychological methods." *arXiv preprint arXiv:2303.13988 1* (2023).
- Yang, Huzheng, James Gee, and Jianbo Shi. "Brain decodes deep nets." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.
- Rathi, Neil, et al. "TopoLM: brain-like spatio-functional organization in a topographic language model." *The Thirteenth International Conference on Learning Representations*.

感谢聆听 !



主页 : <https://ercong21.github.io/>
邮箱 : nie@cis.lmu.de

欢迎关注Munich NLP公众号

关于我们



我们是一群生活在 德国慕尼黑的 NLP 爱好者。我们在这里向大家分享组里的研究成果和感兴趣的论文。我们也欢迎慕尼黑内外的 NLPer 踊跃投稿。

#MunichNLP #LMU #TUM
#NLP



目录

一、个人研究背景和兴趣

二、NLP博士生涯的大模型冲击

三、跨学科与交叉创新重要性

四、交流讨论