

Contents

- Page 1: 221(g) letter
- Page 2: My detailed resume
- Page 3-10: My research advisor's resume
- Page 11: Research and study plan
- Page 12-13: Official transcript
- Page 14: Attended conference and visiting
- Page 15-18: Published paper's abstracts and acceptance notifications



EMBASSY OF THE UNITED STATES OF AMERICA

美国驻华大使馆 - 北京市朝阳区安家楼路 55 号

BEIJING, CHINA – NONIMMIGRANT VISA UNIT / 非移民签证处

<https://china.usembassy-china.org.cn/embassy-consulates/beijing/>

221(g) 行政审理单

This 221(g) letter is valid for exactly ONE
YEAR from the date it is issued.
221(g) 审理单自签发之日起一年内有效。

Dear Applicant: Your application has been refused under Section 221(g) of the Immigration and Nationality Act of 1952. Please be advised that for U.S. visa purposes, including ESTA (<https://esta.cbp.dhs.gov/esta/>), this decision constitutes a denial of a visa. This refusal may be overcome without filing another visa application once you present the required additional evidence. **ALL MATERIALS MUST BE IN ENGLISH.** Additional processing may take four weeks or longer.

根据《移民及国籍法》的 221(g) 条款，您的签证被拒签了。请知悉，对包括通过旅游授权电子系统的旅客，即 ESTA (<https://esta.cbp.dhs.gov/esta/>) 在内的签证申请者来说，根据该条款所作的决定等同于签证被拒。如您能尽快提供补充资料，本次申请将得到进一步审理。您无需重新提交新的申请。所有补充材料必须为英文。行政审理可能需要四周或者更长的时间。

<input checked="" type="checkbox"/>	Detailed CV or resume, including a list of publications 详细的英文个人简历，包括所有出版物的清单	<input checked="" type="checkbox"/>	Complete itinerary, including all meetings, conferences, and visits; include names, addresses, and telephone numbers of your hosts 完整的在美行程，包括所有要参加的会议、谈判和访问， 以及接待方的名称、地址和电话
<input checked="" type="checkbox"/>	Detailed CV or resume of research advisor in the U.S., including email address and a list of publications 在美导师的个人简历，包括电子邮件、出版物清单	<input checked="" type="checkbox"/>	Invitation letter(s) from business, conference, or school, including abstract of paper (if applicable) 商务 / 会议 / 学校 邀请信，包括文章摘要（如适用）
<input checked="" type="checkbox"/>	Research and/or training plan and details of course of study 研究和（或）培训计划以及学习的具体课程	<input checked="" type="checkbox"/>	Current enrollment letter and/or official transcript 当前的注册证明和（或）学校正式成绩单
<input type="checkbox"/>	Proof of SEVIS fee payment 已付 SEVIS 费的证明 (www.fmjfee.com)	<input type="checkbox"/>	Court / Police / Legal documents relating to your situation 由法院 / 警察局 / 律师出具的与您情况相关联的材料
<input type="checkbox"/>	Detailed description of your job / company / equipment for purchase, including end uses and users 关于您职务 / 公司 / 将要购买的设备，包括设备用途及 用户的详细说明	<input type="checkbox"/>	Proof of relationship (invite letter, birth certificate, marriage certificate, copy of U.S. inviter's visa / green card / naturalization record/ U.S. passport) 关系证明（邀请信、出生证、结婚证、美方邀请人的签证 / 绿卡 / 公民证 / 护照的复印件）
<input type="checkbox"/>	Previous passport(s) / visas and/or evidence of U.S. stay and extension 以前的护照 / 签证或在美国延期的证明	<input type="checkbox"/>	OTHER / 其它

Please EMAIL this green form, the requested information, your date of birth, and a Chinese phone number to beijingvisaapp@state.gov with the subject line "LAST NAME, FIRST NAME – PASSPORT NUMBER – DS-160 CONFIRMATION BARCODE". Attachments must be in MS Word, PDF, or JPEG format. If you do not receive an automatic reply email acknowledging receipt of your email, it means we did not receive your email. In this case, please try re-sending your attachments from a different email address.

请将此绿色的表格、所需资料、您的出生日期和中国大陆电话发送电子邮件到 beijingvisaapp@state.gov。邮件标题栏请以如下格式书写“姓，名—护照号码—DS-160 确认页的条形码”邮件附件必须是 MS Word、PDF 或 JPEG 格式。如果您未收到自动回复，这表明我们并没有收到您的邮件。在这种情况下，请您尝试使用不同的邮箱再次发送带有附件的邮件。

Regardless of where you interviewed, please bring this letter with your valid, personal ID and requested items to the Embassy (No. 55 An Jia Lou Road) for an interview 面谈 / to provide fingerprints 提供指纹 / pay relevant visa fees 支付相关签证费. You do not need to make an appointment or pay the application fee again. Please note you will **ONLY** be allowed entry to the Embassy consular section on a Monday, Tuesday, Thursday or Friday during the hours circled below. Please visit <https://china.usembassy-china.org.cn/tag/holiday-schedule/> to confirm the Embassy is open.

请带好本页信函，有效身份证件以及所需补充的英文资料亲自到大使馆(朝阳区安家楼路 55 号)面谈/提供指纹/支付相关签证费。您无需重新预约或再次支付申请费用。请务必于周一，周二，周四或周五的以下标注时间段入馆。请提前上网查看使馆开放时间，我们的网址：<https://china.usembassy-china.org.cn/tag/holiday-schedule/>

8:00am-10:00am (上午 8:00-10:00)

2:00pm-4:00pm (下午 14:00-16:00)

NAME (名字): _____ PASSPORT # (护照号码): _____

Jerry(Jiarui) Zhang

(213) 561-9849 jiaruizhang0613@gmail.com saccharomycetes.github.io/ Los Angeles, CA 90007

EDUCATION

University of Southern California

Ph.D. in Computer Science, Advisor: Willie Neiswanger, GPA: 3.78/4.00

August 2022-Present

Los Angeles, USA

Tsinghua University

Bachelor of Science, Electric Engineering, GPA: 3.64/4.00

August 2018-June 2022

Beijing, China

PUBLICATIONS (* DENOTES EQUAL CONTRIBUTION)

[GOOGLE SCHOLAR](#)

- **J Zhang**, O liu, Y Tianyu, H Jinyi, W Neiswanger. ‘*Euclid: Supercharging Multimodal LLMs with Synthetic High-Fidelity Visual Descriptions*’, In submission.
- Y Jiang*, **J Zhang***, K Sun*, Z Sourati, K Ahrabian, K Ma, F Ilievski, J Pujara. ‘*MARVEL: Multidimensional Abstraction and Reasoning through Visual Evaluation and Learning*’, NeurIPS D&B Track 2024.
- **J Zhang**. ‘*Guided Profile Generation Improves Personalization with LLMs*’, EMNLP Findings 2024.
- K Ahrabian, Z Sourati, K Sun, **J Zhang**, Y Jiang, F Morstatter, J Pujara. ‘*The curious case of nonverbal abstract reasoning with multi-modal large language models*’, COLM 2024.
- **J Zhang**, M Khayatkhoei, P Chhikara, F Ilievski. ‘*Towards Perceiving Small Visual Details in Zero-shot Visual Question Answering with Multimodal LLMs*’, NeurIPS R0-FoMo Workshop, 2023
- **J Zhang**, F Ilievski, K Ma, A Kollaa, J Francis, A Oltramari. ‘*A Study of Situational Reasoning for Traffic Understanding*’, KDD 2023.
- P Chhikara, **J Zhang**, F Ilievski, J Francis, K Ma. ‘*Knowledge-enhanced agents for interactive text games*’, K-Cap, 2023 (Best Student Paper Award)
- **J Zhang**, F Ilievski, K Ma, J Francis, A Oltramari. ‘*A Study of Zero-shot Adaptation with Commonsense Knowledge*’, AKBC 2022.
- X Yi, **J Zhang**, W Li, X Wang, X Xie. ‘*Clickbait Detection via Contrastive Variational Modelling of Text and Label*’, IJCAI 2022.
- **J Zhang***, J Hu*, M Khayatkhoei, F Ilievski, M Sun. ‘*Exploring perceptual limitation of multimodal large language models*’, In submission.

EXPERIENCE

University of Southern California

Graduate Student Researcher

August 2022-Present

Los Angeles, USA

- High-fidelity data training, geometry understanding for multimodal LLMs.
- Exploring and addressing multimodal LLMs’ perceptual limitations on small visual details.
- Utilizing knowledge graph to improve language models. Exploring the application in traffic understanding scenarios.

Amazon Alexa AI

Applied Scientist Intern

May 2023-August 2023

Bellevue, USA

- Personalization enhancement with large language models on both efficacy and efficiency by step-wise self-prompt optimization.

Tsinghua University

Undergraduate Student

December 2020-May 2021

Beijing, China

- Topic-controllable Chinese poetry generation with language models via pseudo-labeling from traditional unsupervised topic models (LDA).

PROGRAMMING

Python, Shell Scripting, \LaTeX

TEACHING

Teaching Assistant for USC DSCI 552

Summer 2024

Teaching Assistant for USC DSCI 552

Fall 2024

CONTACT INFORMATION	216 Powell Hall Los Angeles, CA 90089	neiswang@usc.edu (503) 464-6152
RESEARCH INTERESTS	I develop machine learning methods to perform efficient optimization and experimental design in costly real-world settings, where resources are limited. My work spans topics in active learning, uncertainty quantification, Bayesian decision making, and reinforcement learning. I apply these methods downstream to solve problems in science and engineering, for example in the physical sciences and machine learning systems. I have also worked on distributed algorithms for scalable machine learning, and I develop and maintain software libraries for multilevel optimization, uncertainty quantification, AutoML, and Bayesian optimization.	
CURRENT POSITION	University of Southern California , Los Angeles, CA, United States <i>Assistant Professor, Thomas Lord Department of Computer Science</i> I lead a group that is working at the intersection of machine learning, decision making, generative AI, and AI-for-science.	January 2024 – Present
EDUCATION	Stanford University , Stanford, CA, United States <i>Postdoctoral Scholar, Computer Science Department</i> Working with Stefano Ermon, and affiliated with the StatsML Group, Stanford AI Lab, and SLAC National Accelerator Laboratory.	September 2020 – August 2023
	Carnegie Mellon University , Pittsburgh, PA, United States <i>Ph.D. in Machine Learning</i> Advisor: Eric Xing Thesis: Post-Inference Methods for Scalable Probabilistic Modeling and Sequential Decision Making	September 2012 – August 2019
	Columbia University , New York, NY, United States <i>Bachelor of Science</i> Major: Applied Mathematics, Minor: Computer Science Research Advisors: Chris Wiggins and Frank Wood	September 2008 – May 2012
RESEARCH EXPERIENCE	Stanford University , Stanford, CA, United States <i>Probabilistic Methods, Decision Making, Uncertainty Quantification</i> Working with Stefano Ermon on probabilistic methods for sequential decision making under uncertainty, experimental design, and uncertainty quantification, applied to science and ML systems.	September 2020 – Present
	Carnegie Mellon University , Pittsburgh, PA, United States <i>Scalable Probabilistic Modeling and Sequential Decision Making</i> Worked with Eric Xing to develop scalable inference and optimization algorithms, and with Jeff Schneider and Barnabas Poczos to develop methods for sequential decision making under uncertainty.	September 2012 – 2020
	Columbia University , New York, NY, United States <i>Statistics and Computer Vision Research</i> Worked under the supervision of Frank Wood to develop probabilistic models and inference algorithms for unsupervised detection, tracking, and summarization of objects in videos.	March 2011 – May 2012
	<i>Machine Learning and Cell Biology Research</i> Worked under the supervision of Chris Wiggins to develop software to detect, track, and analyze the motility of T cells for members of the Michael Dustin Laboratory at the NYU Skirball Institute.	October 2010 – July 2012

Conference and Journal Publications

^{*}Denotes equal contribution. [†]Denotes senior authorship.

S. Chitturi, A. Ramdas, Y. Wu, B. Rohr, S. Ermon, J. Dionne, F. Jornada, M. Dunne, C. Tassone[†], W. Neiswanger[†], D. Ratner[†]. “Targeted materials discovery using Bayesian algorithm execution.” *npj Computational Materials*, 2024.

T. Wu*, W. Neiswanger*, H. Zheng*, S. Ermon, J. Leskovec. “Uncertainty Quantification for Forward and Inverse Problems of PDEs via Latent Global Evolution.” *AAAI Conference on Artificial Intelligence (AAAI)*, 2024.

R. Roussel, A. Edelen, T. Boltz, D. Kennedy, Z. Zhang, X. Huang, D. Ratner, A. Garcia, C. Xu, J. Kaiser, A. Eichler, J. Lubsen, N. Isenberg, Y. Gao, N. Kuklev, J. Martinez, B. Mustapha, V. Kain, W. Lin, S. Liuzzo, J. St. John, M. Streeter, R. Lehe, W. Neiswanger. “Bayesian Optimization Algorithms for Accelerator Physics.” *Physical Review Accelerators and Beams (PRAB)*, 2024.

V. Mehta, J. Barr, J. Abbate, M. Boyer, I. Char, W. Neiswanger, E. Kolemen, J. Schneider. “Automated experimental design of safe rampdowns via probabilistic machine learning.” *Nuclear Fusion*, 2024.

S. Miskovich*, W. Neiswanger*, W. Colocco, C. Emma, J. Garrahan, T. Maxwell, C. Mayes, S. Ermon, A. Edelen, D. Ratner “Multipoint-BAX: a new approach for efficiently tuning particle accelerator emittance via virtual objectives.” *Machine Learning: Science and Technology (MLST)*, 2024.

S. Choe, S. Mehta, H. Ahn, W. Neiswanger, P. Xie, E. Strubell, E. Xing. “Making Scalable Meta Learning Practical.” *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

Y. Yuan, C. Chen, Z. Liu, W. Neiswanger, X. Liu. “Importance-aware Co-teaching for Offline Model-based Optimization.” *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

S. Choe, W. Neiswanger, P. Xie, E. Xing. “Betty: An Automatic Differentiation Library for Multilevel Optimization.” *International Conference on Learning Representations (ICLR)*, 2023.

X. Li, V. Mehta, J. Kirschner, I. Char, W. Neiswanger, J. Schneider, A. Krause, I. Bogunovic “Near-optimal Policy Identification in Active Reinforcement Learning.” *International Conference on Learning Representations (ICLR)*, 2023.

B. Boecking, N. Roberts, W. Neiswanger, S. Ermon, F. Sala, A. Dubrawski. “Generative Modeling Helps Weak Supervision (and Vice Versa).” *International Conference on Learning Representations (ICLR)*, 2023.

L. Yu, T. Yu, J. Song, W. Neiswanger, S. Ermon. “Offline Imitation Learning with Suboptimal Demonstrations via Relaxed Distribution Matching.” *AAAI Conference on Artificial Intelligence (AAAI)*, 2023.

W. Neiswanger*, L. Yu*, S. Zhao, C. Meng, S. Ermon. “Generalizing Bayesian Optimization with Decision-theoretic Entropies.” *Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

V. Mehta, I. Char, J. Abbate, R. Conlin, M. Boyer, S. Ermon, J. Schneider, W. Neiswanger[†]. “Exploration via Planning for Information about the Optimal Trajectory.” *Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

Y. Xiao, P. Liang, U. Bhatt, W. Neiswanger, R. Salakhutdinov, L. P. Morency. “Uncertainty Quantification with Pre-trained Language Models: A Large-Scale Empirical Analysis.” *Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2022.

J. Song, L. Yu, W. Neiswanger, S. Ermon. “A General Recipe for Likelihood-free Bayesian Optimization.” *International Conference on Machine Learning (ICML)*, 2022.

- C. Marx, S. Zhao, W. Neiswanger, S. Ermon. “Modular Conformal Calibration.” *International Conference on Machine Learning (ICML)*, 2022.
- V. Mehta, B. Paria, J. Schneider, S. Ermon, W. Neiswanger[†]. “An Experimental Design Perspective on Model-Based Reinforcement Learning.” *International Conference on Learning Representations (ICLR)*, 2022.
- M. Fenstermacher et al. “DIII-D research advancing the physics basis for optimizing the tokamak approach to fusion energy.” *Nuclear Fusion*, 2022.
- C. Meng, E. Liu, W. Neiswanger, J. Song, M. Burke, D. Lobell, S. Ermon. “IS-COUNT: Large-scale Object Counting from Satellite Images with Covariate-based Importance Sampling.” *AAAI Conference on Artificial Intelligence (AAAI)*, 2022.
- A. Qiao, S. Choe, S. Subramanya, W. Neiswanger, Q. Ho, H. Zhang, G. Ganger, E. Xing. “Pollux: Co-adaptive Cluster Scheduling for Goodput-Optimized Deep Learning.” *15th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2021. Recipient of the Jay Lepreau Best Paper Award.
- Y. Chung, W. Neiswanger, I. Char, J. Schneider. “Beyond Pinball Loss: Quantile Methods for Calibrated Uncertainty Quantification.” *Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- A. Narayan, P. Molino, K. Goel, W. Neiswanger, C. Re. “Personalized Benchmarking with the Ludwig Benchmarking Toolkit.” *Datasets and Benchmarks Track, Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- Y. Liu, S. Khandagale, C. White, W. Neiswanger. “Synthetic Benchmarks for Scientific Research in Explainable Machine Learning.” *Datasets and Benchmarks Track, Conference on Neural Information Processing Systems (NeurIPS)*, 2021.
- W. Neiswanger, K. Wang, S. Ermon. “Bayesian Algorithm Execution: Estimating Computable Properties of Black-box Functions Using Mutual Information.” *International Conference on Machine Learning (ICML)*, 2021.
- K. Tran*, W. Neiswanger*, K. Broderick, E. Xing, J. Schneider, Z. Ulissi. “Computational catalyst discovery: Active classification through myopic multiscale sampling.” *The Journal of Chemical Physics*, 2021.
- W. Neiswanger, A. Ramdas. “Uncertainty quantification using martingales for misspecified Gaussian processes.” *32nd International Conference on Algorithmic Learning Theory (ALT)*, 2021.
- B. Boecking, W. Neiswanger, E. Xing, A. Dubrawski. “Interactive Weak Supervision: Learning Useful Heuristics for Data Labeling.” *International Conference on Learning Representations (ICLR)*, 2021.
- V. Mehta, I. Char, W. Neiswanger, Y. Chung, A. Nelson, M. Boyer, E. Kolemen, J. Schneider. “Neural Dynamical Systems.” *60th IEEE Conference on Decision and Control (CDC)*, 2021.
- C. White, W. Neiswanger, Y. Savani. “BANANAS: Bayesian Optimization with Neural Architectures for Neural Architecture Search.” *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.
- C. White, W. Neiswanger, S. Nolen, Y. Savani. “A Study on Encodings for Neural Architecture Search.” *Conference on Neural Information Processing Systems (NeurIPS)*, 2020.
- K. Kandasamy, K. Vysyaraju, W. Neiswanger, B. Paria, C. Collins, J. Schneider, B. Póczos, E. Xing. “Tuning Hyperparameters without Grad Students: Scalable and Robust Bayesian Optimisation with Dragonfly.” *Journal of Machine Learning Research (JMLR)*, 2020.
- K. Tran*, W. Neiswanger*, J. Yoon, E. Xing, Z. Ulissi. “Methods for Comparing Uncertainty Quantifications for Material Property Predictions.” *Machine Learning: Science and Technology (MLST)*, 2020.

- K. Korovina, S. Xu, K. Kandasamy, W. Neiswanger, B. Póczos, J. Schneider, E. Xing. “ChemBO: Bayesian Optimization of Small Organic Molecules with Synthesizable Recommendations.” *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.
- I. Char, Y. Chung, W. Neiswanger, K. Kandasamy, A. Nelson, M. Boyer, E. Kolemen, J. Schneider. “Offline Contextual Bayesian Optimization.” *Conference on Neural Information Processing Systems (NeurIPS)*, 2019.
- K. Kandasamy, W. Neiswanger, R. Zhang, A. Krishnamurthy, J. Schneider, B. Póczos. “Myopic Posterior Sampling for Adaptive Goal Oriented Design of Experiments.” *International Conference on Machine Learning (ICML)*, 2019.
- V. Mayya, E. Judokusumo, E. Abu-Shah, W. Neiswanger, C. Sachar, D. Depoil, L. Kam and M. L. Dustin. “Cutting Edge: Synapse Propensity of Human Memory CD8 T Cells Confers Competitive Advantage over Naive Counterparts.” *The Journal of Immunology*, 2019.
- K. Kandasamy, W. Neiswanger, J. Schneider, B. Póczos, E. Xing. “Neural Architecture Search with Bayesian Optimisation and Optimal Transport.” *Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- V. Mayya, E. Judokusumo, E. Abu Shah, C. Peel, W. Neiswanger, D. Depoil, D. Blair, C. Wiggins, L. Kam, and M. Dustin. “Durable interactions of T cells with T cell receptor stimuli in the absence of a stable immunological synapse.” *Cell Reports*, 2018.
- W. Neiswanger, E. Xing. “Post-Inference Prior Swapping.” *International Conference on Machine Learning (ICML)*, 2017.
- R. Steorts, M. Barnes, W. Neiswanger. “Performance Bounds for Graphical Record Linkage.” *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.
- F. Caron*, W. Neiswanger*, F. Wood, A. Doucet, M. Davy. “Generalized Pólya Urn for Time-Varying Pitman-Yor Processes.” *Journal of Machine Learning Research (JMLR)*, 2017.
- Y. Wang, V. Sadhanala, W. Dai, W. Neiswanger, S. Sra, E. Xing. “Parallel and Distributed Block-Coordinate Frank-Wolfe Algorithms.” *International Conference on Machine Learning (ICML)*, 2016.
- J. Oliva, W. Neiswanger, B. Póczos, J. Schneider, E. Xing. “Fast Function to Function Regression.” *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2015.
- W. Neiswanger, C. Wang, E. Xing. “Asymptotically Exact, Embarrassingly Parallel MCMC.” *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2014.
- W. Neiswanger, C. Wang, Q. Ho, E. Xing. “Modeling Citation Networks using Latent Random Offsets.” *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2014.
- W. Neiswanger, F. Wood, E. Xing. “The Dependent Dirichlet Process Mixture of Objects for Detection-free Tracking and Object Modeling.” *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2014.
- J. Oliva, W. Neiswanger, B. Póczos, J. Schneider, E. Xing. “Fast Distribution to Real Regression.” *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2014.
- V. Mayya*, W. Neiswanger*, R. Medina, C. Wiggins, and M. Dustin. “Integrative analysis of T cell motility from multi-channel microscopy data using TIAM.” *Journal of Immunological Methods*, 2014.
- V. Mayya, W. Neiswanger, D. Blair, C. Wiggins, M. Dustin. “Characterization of synapse-kinapse balance in CD8 T cells.” *The Journal of Immunology*, 2012.

Technical Reports and Workshop Publications

- I. Char, Y. Chung, R. Shah, W. Neiswanger, J. Schneider. “Correlated Trajectory Uncertainty for Adaptive Sequential Decision Making.” *NeurIPS Workshop on Adaptive Experimental Design and Active Learning in the Real World*, 2023.
- V. Mehta, O. Neopane, V. Das, S. Lin, J. Schneider, W. Neiswanger. “Kernelized Offline Contextual Dueling Bandits.” *ICML Workshop on The Many Facets of Preference Learning*, 2023.
- S. Choe, W. Neiswanger, P. Xie, E. Xing. “Betty: An Automatic Differentiation Library for Multilevel Optimization.” *NeurIPS Workshop on Meta-Learning*, 2022.
- R. Tu, N. Roberts, V. Prasad, S. Nayak, P. Jain, F. Sala, G. Ramakrishnan, A. Talwalkar, W. Neiswanger, C. White. “AutoML for Climate Change: A Call to Action.” *NeurIPS Workshop on Tackling Climate Change with Machine Learning*, 2022.
- V. Mehta, I. Char, J. Abbate, R. Conlin, M. Boyer, S. Ermon, J. Schneider, W. Neiswanger[†]. “Sample-efficient Plasma Control by Planning for Optimal Trajectory Information.” *ICML Workshop on Adaptive Experimental Design and Active Learning in the Real World*, 2022.
- Y. Chung, I. Char, H. Guo, J. Schneider, W. Neiswanger[†]. “Uncertainty toolbox: an open-source library for assessing, visualizing, and improving uncertainty quantification.” *ICML Workshop on Uncertainty and Robustness in Deep Learning*, 2021.
- B. Lengerich, W. Neiswanger, E. Lengerich, E. Xing. “Disentangling Increased Testing From Covid-19 Epidemic Spread.” *medRxiv 2020.07.09.20141762*, 2020.
- Y. Chung, I. Char, W. Neiswanger, K. Kandasamy, A. Nelson, M. Boyer, E. Kolemen, J. Schneider. “Offline Contextual Bayesian Optimization for Nuclear Fusion.” *NeurIPS Workshop on Machine Learning and the Physical Sciences*, 2019.
- K. Korovina, S. Xu, K. Kandasamy, W. Neiswanger, B. Póczos, J. Schneider, E. Xing. “ChemBO: Bayesian Optimization of Small Organic Molecules with Synthesizable Recommendations.” *NeurIPS Workshop on Machine Learning and the Physical Sciences*, 2019.
- C. White, W. Neiswanger, Y. Savani. “Deep Uncertainty Estimation for Model-based Neural Architecture Search.” *NeurIPS Workshop on Bayesian Deep Learning*, 2019.
- C. White, W. Neiswanger, Y. Savani. “Neural Architecture Search via Bayesian Optimization with a Neural Network Prior.” *NeurIPS Workshop on Meta-Learning*, 2019.
- W. Neiswanger, X. Liu, E. Xing, “Low Communication Distributed Black Box VI”, *The International Conference on Probabilistic Programming (PROBPROG)*, 2018.
- K. Kandasamy, W. Neiswanger, R. Zhang, A. Krishnamurthy, J. Schneider, B. Póczos. “Sequential Bayesian Design of Experiments via Probabilistic Programming”, *The International Conference on Probabilistic Programming (PROBPROG)*, 2018.
- W. Neiswanger, E. Xing, “Prior Swapping for Data-Independent Inference”, *ICML Workshop on Data-Efficient Machine Learning*, 2016.
- W. Neiswanger, C. Wang, E. Xing. “Embarrassingly Parallel Variational Inference in Nonconjugate Models.” *Tech. Report*. 2015. *arXiv:1510.04163*, 2015.
- W. Neiswanger, C. Wang, E. Xing, “Embarrassingly Parallel Variational Inference.” *NeurIPS Workshop on Advances in Variational Inference*, 2014.

Y. Wang, V. Sadhanala, W. Dai, W. Neiswanger, S. Sra, E. Xing, “Asynchronous Parallel Block-Coordinate Frank-Wolfe”, *NeurIPS Workshop on Optimization for Machine Learning*, 2014.

W. Neiswanger, C. Wang, E. Xing, “Embarrassingly Parallel MCMC via Density Product Estimation”, *NIPS Workshop on Randomized Methods for Machine Learning*, 2013.

W. Neiswanger and F. Wood, “Unsupervised Detection and Tracking of Multiple Objects with Dependent Dirichlet Process Mixtures”, *New York Academy of Sciences, Machine Learning Symposium*, New York, NY. October 21, 2011.

AWARDS

Jay Lepreau Best Paper Award, OSDI, 2021

ICML Reviewer Award, 2020

AWS Machine Learning Research Award, 2019

Pittsburgh Filmmakers Spring Mix, Audience Choice Award, 2017

ICML Reviewer Award, 2015

Columbia University Department of Applied Mathematics Faculty Award, 2012

KAUST International Research Competition, 1st Place, 2012

CUSP Summer Enhancement Fellowship Winner 2009, 2011

Columbia C.P. Davis Scholar, 2008

Intel ISEF Grand Prize Award Winner, 4th Place 2006, 3rd Place 2007, 4th Place 2008

INVITED TALKS

AI-Guided Experiments with Bayesian Algorithm Execution. Toyota Research Institute (TRI), October 2022.

AI-Guided Experiments for ML Systems, AutoML, and Science. 1st CASL Project Workshop, October 2022.

Bayesian Algorithm Execution for Tuning Particle Accelerator Emittance with Partial Measurements. General Assembly of the Helmholtz International Laboratory (HIR³X), September 2022.

BAX, InfoBAX, and Applications to Experimental Design and Reinforcement Learning. Oxford University Deep-Prob Seminar, February 2022.

Going Beyond Global Optima with Bayesian Algorithm Execution. AutoML Seminars, European Lab for Learning and Intelligent Systems (ELLIS), December 2021.

AI Meets Biology Panel. State-Of-The-Art Conference, December 2021.

InfoBAX: Estimating Black-box Functions Using Mutual Information. London Data Science Journal Club, November 2021.

Going Beyond Global Optima with Bayesian Algorithm Execution. Saint Petersburg State University, Industrial Mathematics Seminar, September 2021.

Uncertainty quantification in machine learning with applications to sequential decision making. Korea University, Machine learning and Policy Studies International Workshop, September 2021.

Going Beyond Global Optima with Bayesian Algorithm Execution. University College London (UCL), Statistical Machine Learning Group, July 2021.

Going Beyond Global Optima with Bayesian Algorithm Execution. LBNL Camera Workshop on Autonomous Discovery in Science and Engineering, April 2021.

Going Beyond Global Optima with Bayesian Algorithm Execution. AI Seminar, SLAC National Accelerator Laboratory, April 2021.

Working with Scientists to Develop Uncertainty Models for Improved Sequential Decision Making. SIAM CSE21 Minisymposium, March 2021.

Probabilistic Programming for Sequential Decision Making Under Uncertainty. Stanford University, March 2020.

Machine Learning Bandits. Carnegie Mellon University TechNights Program, March 2018.

Post-Inference Prior Swapping. Machine Learning Lunch Seminar. Carnegie Mellon University, November 2017.

Variational Bayesian Inference and Mean Field Approximations. Guest lecture, Probabilistic Graphical Models Class. Carnegie Mellon University, March 2017.

Embarrassingly Parallel MCMC. Machine Learning Lunch Seminar. Carnegie Mellon University, October 2014.

Asymptotically Exact, Embarrassingly Parallel MCMC. Oxford Computational Statistics / Statistical Machine Learning Reading Group. Oxford University, July 2014.

Asymptotically Exact, Embarrassingly Parallel MCMC. Department of Information Engineering Tea Talk. Oxford University, July 2014.

Unsupervised Detection and Tracking of Multiple Objects with Dependent Dirichlet Process Mixtures. Tutorial for Statistics Department Data Mining Class. Columbia University, January 2012.

TEACHING	<i>Guest Lecture for Probabilistic Graphical Models Class at CMU</i> Gave a guest lecture on variational inference and mean field approximations.	Spring 2017
	<i>Two Guest Lectures for Probabilistic Graphical Models Class at CMU</i> Gave two lectures on approximate inference algorithms and probabilistic latent variable models.	Spring 2015
	<i>Teaching Assistant for CMU 10-601: Introduction to Machine Learning</i> Led recitation lectures, mentored project groups, designed tests and homework assignments.	Spring 2015
	<i>Teaching Assistant for CMU 10-708: Probabilistic Graphical Models</i> Led recitation lectures, mentored project groups, designed tests and homework assignments.	Spring 2014
	<i>Guest Lecture for Data Mining and Machine Learning Class at Columbia University</i> Gave a guest lecture on nonparametric Bayesian models for computer vision applications.	Spring 2012
ACTIVITIES AND PROFESSIONAL SERVICE	<i>Workshop Organizer</i> Co-organized the NeurIPS 2023 Workshop on Adaptive Experimental Design and Active Learning in the Real World.	December 2023
	<i>Workshop Organizer</i> Co-organized the ICML 2023 Workshop on Differentiable Almost Everything: Differentiable Relaxations, Algorithms, Operators, and Simulators.	July 2023
	<i>Workshop Organizer</i> Co-organized the ICML 2022 Workshop on Adaptive Experimental Design and Active Learning in the Real World.	July 2022
	<i>Workshop Organizer</i> Co-organized the ICML 2021 Workshop on Machine Learning for Data.	July 2021
	<i>CASL Project</i> Co-founded and organized the CASL Project for composable, automatic, and scalable ML systems.	September 2020 – Present
	<i>Workshop Organizer</i> Co-organized the ICML 2020 Workshop on Real World Experiment Design and Active Learning.	July 2020

<i>Reviewer</i>	June 2013 – Present
NeurIPS, ICML, JMLR, ICLR, AISTATS, UAI, AAAI, IJCAI, CVPR, TKDE, JSIG, AISM, Statistics and Computing, Digital Signal Processing, PLOS ONE, JASA, Biometrika, Journal of the Royal Statistical Society: Series B.	
<i>Workshop Organizer</i>	December 2019
Co-organized the NeurIPS 2019 Workshop on Learning with Rich Experience (LIRE).	
<i>CMU AI+ Club</i>	January 2018 – August 2020
Co-founded Carnegie Mellon's AI+ club. Helped organize and host events.	
<i>CMU Machine Learning Lunch Seminar</i>	September 2013 – August 2019
Organized Carnegie Mellon's Machine Learning Lunch weekly seminar.	
<i>Columbia University Beginner Machine Learning Reading Group</i>	June 2011 – May 2012
Founder, participant, and presenter. Helped organize and conduct meetings.	
<i>Society for Industrial and Applied Mathematics (SIAM)</i>	May 2011 – May 2012
Member of SIAM and board member of the Columbia University SIAM student chapter.	
<i>Tutoring</i>	June 2010 – May 2012
Tutored high school students in mathematics (calculus, precalculus, trigonometry).	
PROGRAMMING	Python, R, SQL, Shell Scripting, \LaTeX .
SOFTWARE AND SYSTEMS	<p><i>LLM360</i> https://www.llm360.ai/ Fully-transparent open-source large language model research and development.</p> <p><i>Uncertainty Toolbox</i> https://uncertainty-toolbox.github.io A Python toolbox for predictive uncertainty quantification, calibration, metrics, and visualization.</p> <p><i>Betty</i> https://github.com/leopard-ai/betty An automatic differentiation library for generalized meta-learning and multilevel optimization.</p> <p><i>AdaptDL</i> https://github.com/petuum/adaptdl Resource-adaptive cluster scheduler for deep learning training.</p> <p><i>TorchUQ</i> https://github.com/TorchUQ/torchuq A library for uncertainty quantification based on PyTorch.</p> <p><i>Naszilla</i> https://github.com/naszilla/naszilla A Python library for neural architecture search (NAS).</p> <p><i>Dragonfly</i> https://github.com/dragonfly/dragonfly An open source Python library for scalable Bayesian optimisation.</p>

Research and Study Plan

I have finished 2 years of my Ph.D. program, the list of courses I studied are included in my current transcript, including topics about natural language processing, ethics of artificial intelligence and algorithms. I have finished most of the course works and will focus more on research projects in the following years.

My research is focus on natural language processing and multimodal learning. Recently, I study the capability and properties of multimodal large language model, the artificial intelligence model that can read, see and tell. My goal is to improve these models' understanding about the physical world and help them become useful assistants for human.

Year 3:

- Serve as a TA for Computer Science classes.
- Learn related papers and study practical cases, continue my research in multimodal learning, finish 1 to 2 papers in this research field.

Year 4:

- Serve as research assistant for my advisor, continue my study and research in multimodal learning, and continue publishing high-influence papers.

Year 5:

- Finish dissertation and papers for the job market; write papers as much as I can and try to get them published
- Pass final examination (oral defense of the dissertation)
- Get the Ph.D. degree and graduate from USC

UNIVERSITY OF SOUTHERN CALIFORNIA
OFFICIAL ACADEMIC TRANSCRIPT

OFFICE OF THE REGISTRAR
LOS ANGELES, CA 90089-0912
(213) 740-7445

RELEASE OF THIS RECORD OR DISCLOSURE OF ITS
CONTENTS TO ANY THIRD PARTY WITHOUT WRITTEN
CONSENT OF THE STUDENT IS PROHIBITED

STUDENT NAME	STUDENT NUMBER	DATE	PAGE
Zhang, Jiarui	3709-3529-97	12-17-2024	2 of 2

NOTE: PHOTOCOPIES ARE NOT TO BE CONSIDERED OFFICIAL TRANSCRIPTS. THE REGISTRAR'S SEAL AND SIGNATURE
APPEAR ON THE FIRST PAGE.

Spring Semester 2023 (01-09-2023 to 05-12-2023)

CSCI-544	A	4.0	Applied Natural Language Processing
ALI-244	CR	2.0	Academic and Professional Speaking Skills III
CSCI-697	CR	1.0	Seminar in Computer Science Research
CSCI-572	A	4.0	Information Retrieval and Web Search Engines

Term Units Attempted	Term Units Earned	Term GPA	Term Grade Points	Term GPA
11.0	11.0	8.0	32.00	4.00

Summer Semester 2023 (05-17-2023 to 08-08-2023)

ENGR-596	CR	1.0	Internship in Engineering
----------	----	-----	---------------------------

Term Units Attempted	Term Units Earned	Term GPA	Term Grade Points	Term GPA
1.0	1.0	0	0	0.00

Fall Semester 2023 (08-21-2023 to 12-13-2023)

CSCI-699	A-	4.0	Special Topics (Ethics in Natural Language Processing)
CSCI-654	A	4.0	Crafting a Research Agenda
ALI-245	CR	2.0	Academic and Professional Writing III

Term Units Attempted	Term Units Earned	Term GPA	Term Grade Points	Term GPA
10.0	10.0	8.0	30.80	3.85

Spring Semester 2024 (01-08-2024 to 05-10-2024)

CSCI-697	CR	1.0	Seminar in Computer Science Research
CSCI-790	CR	2.0	Research
CSCI-670	A-	4.0	Advanced Analysis of Algorithms

Term Units Attempted	Term Units Earned	Term GPA	Term Grade Points	Term GPA
7.0	7.0	4.0	14.80	3.70

End of Transcript

Conference attended so far:

- AAAI 2023 (<https://aaai-23.aaai.org/>)
 - Location: Washington D.C.
 - Contact: aaai23@aaai.org
- KDD 2023 (<https://kdd.org/kdd2023/index.html>)
 - Location: Long Beach, CA
 - Contact: KDD23-general@acm.org
- NeurIPS 2023 (<https://neurips.cc/Conferences/2023>)
 - Location: New Orleans
 - Contact: <https://neurips.cc/Help/Contact>
- EMNLP 2024
 - Location: Miami, Florida
 - Contact: emnlp2024-programchairs@googlegroups.com

I will attend more conferences in the coming years to present my published papers. Typically, these conferences will happen all around the world so I need a longer term US visa to attend the conference and return to my school during the rest of my PhD.

I have not officially visited any organizations and do not have a plan to do so.

Paper acceptance and abstract (AAAI 2023 workshop paper):

Paper ID	Title	Files	Status	Actions
28	<p>Utilizing Background Knowledge for Robust Reasoning over Traffic Situations</p> <p>Hide abstract</p> <p>Understanding novel situations in the traffic domain requires a complex association of domain-specific and causal commonsense knowledge. Prior work has generally focused on situational reasoning in a general domain or visual processing of traffic situations.</p> <p>In this paper, we focus on an evaluation of robust reasoning over traffic situations expressed in natural language. We put forward two text-based multiple-choice question answering (MCQA) sets: BDD-QA for evaluating causal reasoning in the traffic domain and HDT-QA for measuring the possession of domain knowledge akin to human driving license tests.</p> <p>We adapt three knowledge-driven approaches for zero-shot QA over traffic situations, based on prior methods for natural language inference, commonsense models with knowledge graph self-supervision, and dense retriever-based models.</p> <p>Our experiments show that language models trained with inference information and commonsense knowledge perform well on the BDD-QA dataset, while efficient knowledge extraction leads to the best performance on the HDT-QA dataset.</p>	<p>Submission files:</p> <p>① Utilizing Background Knowledge for Robust Reasoning over Traffic Situations.pdf</p> <p>Camera Ready Submission files:</p> <p>④ Utilizing Background Knowledge for Robust Reasoning over Traffic Situations (4).pdf</p>	<p>Accept (poster)</p> <p>Reviews</p> <p>Meta-Reviews</p>	<p>Email: ✉ Email</p> <p>Meta-Reviewer</p>

Paper acceptance, abstract and decision notification (KDD 2023 paper):

A Study of Situational Reasoning for Traffic Understanding



Anonymous

Published: 17 May 2023, Last Modified: 17 May 2023 KDD 2023 Research Track Readers: Research Track, Paper1113 Area Chairs, Paper1113 Reviewers, Paper1113

Authors Show Bibtex Show Revisions

Keywords: traffic understanding, zero-shot evaluation, language models, question answering

TL;DR: This paper studies the ability of diverse knowledge-enhanced language models to reason over traffic situations in three novel reasoning benchmarks.

Abstract: Intelligent Traffic Monitoring (ITMo) technologies hold the potential for improving road safety/security and for enabling smart city infrastructure. Understanding traffic situations requires a complex fusion of perceptual information with domain-specific and causal commonsense knowledge. Whereas prior work has provided benchmarks and methods for traffic monitoring, it remains unclear whether models can effectively align these information sources and reason in novel scenarios. To address this assessment gap, we devise three novel text-based tasks for situational reasoning in the traffic domain: i) BDD-QA, which evaluates the ability of Language Models (LMs) to perform situational decision-making, ii) TV-QA, which assesses LMs' abilities to reason about complex event causality, and iii) HDT-QA, which evaluates the ability of models to solve human driving exams.

We adopt four knowledge-enhanced methods that have shown generalization capability across language reasoning tasks in prior work, based on natural language inference, commonsense knowledge-graph self-supervision, multi-QA joint training, and dense retrieval of domain information. We associate each method with a relevant knowledge source, including knowledge graphs, relevant benchmarks, and driving manuals.

In extensive experiments, we benchmark various knowledge-aware methods against the three datasets, under zero-shot evaluation; we provide in-depth analyses of model performance on data partitions and examine model predictions categorically, to yield useful insights on traffic understanding, given different background knowledge and reasoning strategies.

Subject Areas: Classification, Deep Learning & Its Applications, Few-shot Learning and Transfer Learning, Knowledge and Ontologies

Submission Guidelines: Yes

Revealed to Jiarui Zhang, Filip Ilievski, Kaixin Ma, kollaa@isi.edu, Jonathan Francis, Alessandro Oltramari

Published: 17 May 2023, Last Modified: 03 Feb 2023 KDD 2023 Research Track

Authors: Jiarui Zhang, Filip Ilievski, Kaixin Ma, Aravinda Kollaa, Jonathan Francis, Alessandro Oltramari

Student Author: Yes

Serve As Reviewer: name: Filip Ilievski

email: ilievski@isi.edu

homepage: <http://www.ilievski.info/>

Add Withdraw

Reply Type: all Author: everybody Visible To: all readers Hidden From: nobody

8 Replies

[–] Paper Decision

KDD 2023 Conference Research Track Program Chairs

11 May 2023, 22:05 (modified: 12 May 2023, 22:10) KDD 2023 Conference Research Track Paper1113 Decision Readers: Program Chairs, Paper1113

Authors Show Revisions

Decision: Accept

Comment: The paper presents a set of novel text-based tasks for situational reasoning in the traffic domain, which are well motivated and designed.

The contracted datasets can be useful for the research community. The paper is well written where the tasks, approaches, and the experimental evaluation are clearly described. The main limitation is that the paper only considers language tasks without other modalities that are very common in the traffic domain.

[KDD 2023 Research Track] Decision notification for your submission 1113: A Study of Situational Reasoning for Traffic Understanding [Inbox](#)



OpenReview <noreply@openreview.net>

to me ▾

May 17, 2023, 6:37 AM



Dear Jiarui Zhang,

Thank you for submitting your paper, A Study of Situational Reasoning for Traffic Understanding, to KDD 2023 Research Track. We are delighted to inform you that your submission has been accepted. Congratulations!

You can find the final reviews for your paper on the submission page in OpenReview at: <https://openreview.net/forum?id=xcligDlyZQx>

This year, the Research track received a total of 1416 submissions, of which 313 were accepted.

We did our best to obtain at least three independent reviews for every submission. For each submission, the reviews and author responses were evaluated by a dedicated Senior Program Committee Member, who moderated the discussions among the reviewers and provided a summary of the discussion and independent evaluation when needed. We hope that the reviews will be helpful in improving your paper for the camera-ready version.

Please note:

– Camera ready deadline for all the papers is June 4, 2023. ACM has updated [the author policy](#). More detailed instructions on camera ready will be sent later. For any camera-ready paper issues, please contact the Proceedings Chairs: KDD23-proceedings-chairs@acm.org

– Note also that as mentioned in the CFP, no changes to the author list is allowed.

– Due to the site capacity constraints, we will have talks of different lengths, such as regular talks and lightning talks; the length of your talk will be sent later. Every paper will be presented as posters in addition to talks.

– This year KDD is planned to be a fully in-person conference. The registration website is at [the link] (<https://cvcent.me/zykkaV>). The early bird deadline is June 26, 2023. At least one regular registration is required for each paper.

– Please note [ACM's policy against harassment at ACM activities](#).

– Any other items such as video/slideslive preparation format and uploading deadline; student travel award application; PhD forum; poster instructions.

Congratulations again and thank you for submitting your work to the Research track!

Leman Akoglu, Dimitrios Gunopulos, Xifeng Yan

KDD 2023 Research Track Co-Chairs

KDD23-pc-chairs@acm.org

Paper acceptance and abstract (NeurIPS 2023 workshop paper):

Visual Cropping Improves Zero-Shot Question Answering of Multimodal Large Language Models



Jiarui Zhang, Mahyar Khayatkhoei, Prateek Chhikara, Filip Ilievski

Published: 02 Nov 2023, Last Modified: 13 Dec 2023 R0-FoMo Poster Everyone Revisions BibTeX

Keywords: multimodal large language model, visual question answering, visual cropping, visual details, BLIP-2

TL;DR: We qualitatively and quantitatively show the limitation of two state-of-the-art multimodal LLMs in perceiving small visual details for zero-shot visual question answering, and then propose three visual cropping methods to mitigate this limitation.

Abstract:

Multimodal Large Language Models (LLMs) have recently achieved promising zero-shot accuracy on visual question answering (VQA) – a fundamental task affecting various downstream applications and domains. Given the great potential for the broad use of these models, it is important to investigate their limitations in dealing with different image and question properties. In this work, we investigate whether multimodal LLMs can perceive small details as well as large details in images. In particular, we show that their zero-shot accuracy in answering visual questions is very sensitive to the size of the visual subject of the question, declining up to 46 with size. Furthermore, we show that this effect is causal by observing that human visual cropping can significantly mitigate their sensitivity to size. Inspired by the usefulness of human cropping, we then propose three automatic visual cropping methods as inference time mechanisms to improve the zero-shot performance of multimodal LLMs. We study their effectiveness on four popular VQA datasets, and a subset of the VQAv2 dataset tailored towards fine visual details. Our findings suggest that multimodal LLMs should be used with caution in detail-sensitive VQA applications, and that visual cropping is a promising direction to improve their zero-shot performance.

Submission Number: 109

The screenshot shows a search interface with various filters: 'Filter by reply type...', 'Filter by author...', 'Search keywords...', 'Sort: Newest First'. Below the filters is a list of replies with titles like 'Everyone', 'Program Chairs', 'Submission109 Area...', 'Submission109...', 'Submission109 Authors', 'Submission109...', 'Submission109...', 'Submission109...', 'Submission109...'. At the bottom right, it says '5 / 5 replies shown'.

Add: Withdrawal

Paper Decision

Decision by Program Chairs 28 Oct 2023, 03:42 (modified: 28 Oct 2023, 03:47) Program Chairs, Area Chairs, Reviewers, Authors Revisions

Decision: Accept (Poster)

Meta Review of Submission109 by Area Chair mBeM

Meta Review by Area Chair mBeM 26 Oct 2023, 06:19 (modified: 28 Oct 2023, 03:45) Area Chairs, Authors, Reviewers Submitted, Program Chairs Revisions

Metareview:

Reviewers commend the paper's insightful and original contributions while expressing shared concerns about information loss and cropping issues. The paper can hold promise for the broader research community.

Recommendation: Accept (Poster)

Confidence: 5: The area chair is absolutely certain

Paper acceptance and abstract (EMNLP 2024 paper):

Guided Profile Generation Improves Personalization with Large Language Models



Jiarui Zhang

Published: 19 Sept 2024, Last Modified: 03 Oct 2024 EMNLP 2024 Findings Conference, Area Chairs, Reviewers, Publication Chairs, Authors Revisions BibTeX CC BY 4.0

Keywords: personalization, large language models, guided generation

TL;DR: We propose guided profile generation to enhance personalization for large language models and evaluate its effectiveness on three popular personalization tasks.

Track: NLP Applications

Abstract:

In modern commercial systems, including Recommendation, Ranking, and E-Commerce platforms, there is a trend towards improving customer experiences by incorporating Personalization context as input into Large Language Models (LLM). However, LLMs often struggle to effectively parse and utilize sparse and complex personal context without additional processing or contextual enrichment, underscoring the need for more sophisticated context understanding mechanisms. In this work, we propose Guided Profile Generation (GPG), a general method designed to generate personal profiles in natural language. As is observed, intermediate guided profile generation enables LLMs to summarize, and extract the important, distinctive features from the personal context into concise, descriptive sentences, precisely tailoring their generation more closely to an individual's unique habits and preferences. Our experimental results show that GPG improves LLM's personalization ability across different tasks, for example, it increases 37% accuracy in predicting personal preference compared to directly feeding the LLMs with raw personal context.

Paper Link: [/forum?id=upUpGrwFPq](#)

Response To Metareview:

We are glad that AC find our work effective, our experiment to be substantial and our visual examples to be useful. Regarding the efficiency of our framework, please refer to our general response 2 and we will include it in our final version. For the second question, please refer our response to reviewer tPKG: in short, raw preference data is highly unstructured, making a simple semantic similarity search insufficient for such task. The poor performance of direct feeding in Table 1 further support this and demonstrate the effectiveness of generating intermediate personal context digestion and a guided personal profile.

Format Check: Finished ACLPUBcheck

Copyright PDF: pdf

Handbook: I prefer a digital conference handbook

Final Check: I understand that no further revisions can be requested after the deadline.

Paper Type: long

Submission Number: 792

Filter by reply type...	Filter by author...	Search keywords...	Sort: Newest First				-	=	=	
Everyone	Program Chairs	Submission792 Authors		1 / 1 reply shown						

Paper Decision

Decision by Program Chairs 19 Sept 2024, 10:16 (modified: 20 Sept 2024, 23:15) Program Chairs, Authors Revisions

Decision: Accept (Findings)