

圆桌论道 | Neurips 2024 Large Language Model 1022篇相关论文分类整理

作者: RLCN公众号

二维码:

**AI for science****DePLM: Denoising Protein Language Models for Property Optimization**

Zeyuan Wang · Keyan Ding · Ming Qin · Xiaotong Li · Xiang Zhuang · Yu Zhao · Jianhua Yao · Qiang Zhang · Huajun Chen

MutaPLM: Protein Language Modeling for Mutation Explanation and Engineering

Yizhen Luo · Zikun Nie · Massimo Hong · Suyuan Zhao · Hao Zhou · Zaiqing Nie

PhyloGen: Language Model-Enhanced Phylogenetic Inference via Graph Structure Generation

ChenRui Duan · Zelin Zang · Siyuan Li · Yongjie Xu · Stan Z. Li

Generative Hierarchical Materials Search

Sherry Yang · Simon Batzner · Ruiqi Gao · Muratahan Aykol · Alexander Gaunt · Brendan C McMorro · Danilo Jimenez Rezende · Dale Schuurmans · Igor Mordatch · Ekin Dogus Cubuk

MMSite: A Multi-modal Framework for the Identification of Active Sites in Proteins

Song Ouyang · Huiyu Cai · Yong Luo · Kehua Su · Lefei Zhang · Bo Du

Invariant Tokenization of Crystalline Materials for Language Model Enabled Generation

Keqiang Yan · Xiner Li · Hongyi Ling · Kenna Ashen · Carl Edwards · Raymundo Arroyave · Marinka Zitnik · Heng Ji · Xiaofeng Qian · Xiaoning Qian · Shuiwang Ji

Approximating mutual information of high-dimensional variables using learned representations

Gokul Gowri · Xiaokang Lun · Allon Klein · Peng Yin

Multi-Scale Representation Learning for Protein Fitness Prediction

Zuobai Zhang · Pascal Notin · Yining Huang · Aurelie Lozano · Vijil Chenthamarakshan · Debora Marks · Payel Das · Jian Tang

Bridge-IF: Learning Inverse Protein Folding with Markov Bridges

Yiheng Zhu · Jialu Wu · Qiuyi Li · Jiahuan Yan · Mingze Yin · Wei Wu · Mingyang Li · Jieping Ye · Zheng Wang · Jian Wu

SauLLM-54B & SauLLM-141B: Scaling Up Domain Adaptation for the Legal Domain

Pierre Colombo · Telmo Pessoa Pires · Malik Boudiaf · Rui Melo · Gabriel Hautreux · Etienne Malaboeuf · Johanne Charpentier · Dominic Culver · Michael Desa

Sequence-Augmented SE(3)-Flow Matching For Conditional Protein Generation

Guillaume Huguet · James Vuckovic · Kilian FATRAS · Eric Thibodeau-Laufer · Pablo Lemos · Riashat Islam · Chenghao Liu · Jarrod Rector-Brooks · Tara Akhound-Sadegh · Michael Bronstein · Alexander Tong · Joey Bose

Proximal Causal Inference With Text Data

Jacob Chen · Rohit Bhattacharya · Katherine Keith

ProSST: Protein Language Modeling with Quantized Structure and Disentangled Attention

Mingchen Li · Yang Tan · Xinzhu Ma · Bozitao Zhong · Huiqun Yu · Ziyi Zhou · Wanli Ouyang · Bingxin Zhou · Pan Tan · Liang Hong

Can LLMs Implicitly Learn Numeric Parameter Constraints in Data Science APIs?

Yinlin Deng · Chungqiu Steven Xia · Zhezhen Cao · Meiziniu Li · LINGMING ZHANG

HLM-Cite: Hybrid Language Model Workflow for Text-based Scientific Citation Prediction

Qian Yue Hao · Jingyang Fan · Fengli Xu · Jian Yuan · Yong Li

MACM: Utilizing a Multi-Agent System for Condition Mining in Solving Complex Mathematical Problems

Bin Lei · Yi Zhang · Shan Zuo · Ali Payani · Caiwen Ding

Ultrafast classical phylogenetic method beats large protein language models on variant effect prediction

Sebastian Prillo · Wilson Wu · Yun Song

Hybrid Generative AI for De Novo Design of Co-Crystals with Enhanced Tabletability

Nina Gubina · Andrei Dmitrenko · Gleb Solovov · Lyubov Yamshchikova · Oleg Petrov · Ivan Lebedev · Nikita Serov · Grigori Kirgizov · Nikolay Nikitin · Vladimir Vinogradov

Learning Complete Protein Representation by Dynamically Coupling of Sequence and Structure

Bozhen Hu · Cheng Tan · Jun Xia · Yue Liu · Lirong Wu · Jiangbin Zheng · Yongjie Xu · Yufei Huang · Stan Z. Li

Learning Formal Mathematics From Intrinsic Motivation

Gabriel Poesia · David Broman · Nick Haber · Noah Goodman

Training Compute-Optimal Protein Language Models

Xingyi Cheng · Bo Chen · Pan Li · Jing Gong · Jie Tang · Le Song

Enhancing vision-language models for medical imaging: bridging the 3D gap with innovative slice selection

Yuli Wang · Peng jian · Yuwei Dai · Craig Jones · Haris Sair · Jinglai Shen · Nicolas Loizou · jing wu · Wen-Chi Hsu · Maliha Imami · Zhicheng Jiao · Paul Zhang · Harrison Bai

BIOSCAN-5M: A Multimodal Dataset for Insect Biodiversity

Zahra Gharaee · Scott C. Lowe · ZeMing Gong · Pablo Millan Arias · Nicholas Pellegrino · Austin T. Wang · Joakim Bruslund Haurum · Iuliia Eyriay · Lila Kari · Dirk Steinke · Graham Taylor · Paul Fieguth · Angel Chang
fMRI predictors based on language models of increasing complexity recover brain left lateralization
Laurent Bonnasse-Gahot · Christophe Pallier
AI Hallucination
Automated Multi-level Preference for MLLMs
Mengxi Zhang · Wenhao Wu · Yu Lu · YuXin Song · KANG RONG · Huanjin Yao · Jianbo Zhao · Fanglong Liu · Haocheng Feng · Jingdong Wang · Yifan Sun
Toward a Stable, Fair, and Comprehensive Evaluation of Object Hallucination in Large Vision-Language Models
Hongliang Wei · Xingtao Wang · Xianqi Zhang · Xiaopeng Fan · Debin Zhao
Perception of Knowledge Boundary for Large Language Models through Semi-open-ended Question Answering
Zhihua Wen · Zhiliang Tian · Zexin Jian · Zhen Huang · Pei Ke · Yifu Gao · Minlie Huang · Dongsheng Li
To Believe or Not to Believe Your LLM: Iterative Prompting for Estimating Epistemic Uncertainty
Yasin Abbasi Yadkori · Ilja Kuzborskij · András György · Csaba Szepesvari
LLM-Check: Investigating Detection of Hallucinations in Large Language Models
Gaurang Sriramanan · Siddhant Bharti · Vinu Sankar Sadasivan · Shoumik Saha · Priyatham Kattakinda · Soheil Feizi
Mitigating Object Hallucination via Concentric Causal Attention
Yun Xing · Yiheng Li · Ivan Laptev · Shijian Lu
The Factorization Curse: Which Tokens You Predict Underlie the Reversal Curse and More
Ouail Kitouni · Niklas S Nolte · Adina Williams · Michael Rabbat · Diane Bouchacourt · Mark Ibrahim
Detecting Bugs with Substantial Monetary Consequences by LLM and Rule-based Reasoning
Brian Zhang · Zhuo Zhang
Alleviating Hallucinations in Large Vision-Language Models through Hallucination-Induced Optimization
Beitao Chen · Xinyu Lyu · Lianli Gao · Hengtao Shen · Jingkuan Song
Graph-based Uncertainty Metrics for Long-form Language Model Generations
Mingjian Jiang · Yangjun Ruan · Prasanna Sattigeri · Salim Roukos · Tatsunori Hashimoto
Reranking Laws for Language Generation: A Communication-Theoretic Perspective
António Farinhas · Xiaocheng Li · André Martins
Calibrated Self-Rewarding Vision Language Models
Yiyang Zhou · Zhiyuan Fan · Dongjie Cheng · Sihan Yang · Zhaorun Chen · Chenhang Cui · Xiyao Wang · Yun Li · Linjun Zhang · Huaxiu Yao
HaloScope: Harnessing Unlabeled LLM Generations for Hallucination Detection
Xuefeng Du · Chaowei Xiao · Sharon Li
Allegator: Alleviating Attention Bias for Visual-Informed Text Generation
MISO CHOI · Jinyoung Kim · Minseo Yoon · Ji Soo Lee · Hyunwoo Kim
Long-form factuality in large language models
Jerry Wei · Chengrun Yang · Xinying Song · Yifeng Lu · Nathan Hu · Jie Huang · Dustin Tran · Daiyi Peng · Ruibo Liu · Da Huang · Cosmo Du · Quoc V Le
Improving Sparse Decomposition of Language Model Activations with Gated Sparse Autoencoders
Senthooran Rajamanoharan · Arthur Conmy · Lewis Smith · Tom Lieberum · Vikrant Varma · Janos Kramar · Rohin Shah · Neel Nanda
ANAH-v2: Scaling Analytical Hallucination Annotation of Large Language Models
Yuzhe Gu · Ziwei Ji · Wenwei Zhang · Chengqi Lyu · Dahua Lin · Kai Chen
IRCAN: Mitigating Knowledge Conflicts in LLM Generation via Identifying and Reweighting Context-Aware Neurons
Dan Shi · Renren Jin · Tianhao Shen · Weilong Dong · Xinwei Wu · Deyi Xiong
SLED: Self Logits Evolution Decoding for Improving Factuality in Large Language Models
Jianyi Zhang · Da-Cheng Juan · Cyrus Rashtchian · Chun-Sung Ferng · Heinrich Jiang · Yiran Chen
Kernel Language Entropy: Fine-grained Uncertainty Quantification for LLMs from Semantic Similarities
Alexander Nikitin · Jannik Kossen · Yarin Gal · Pekka Marttinen
Many-shot Jailbreaking
Cem Anil · Esin DURMUS · Nina Panickssery · Mrinank Sharma · Joe Benton · Sandipan Kundu · Joshua Batson · Meg Tong · Jesse Mu · Daniel Ford · Francesco Mosconi · Rajashree Agrawal · Rylan Schaeffer · Naomi Bashkansky · Samuel Svenningsen · Mike Lambert · Ansh Radhakrishnan · Carson Denison · Evan Hubinger · Yuntao Bai · Trenton Bricken · Timothy Maxwell · Nicholas Schiefer · James Sully · Alex Tamkin · Tamera Lanham · Karina Nguyen · Tomek Korbak · Jared Kaplan · Deep Ganguli · Samuel Bowman · Ethan Perez · Roger Grosse · David Duvenaud
Multi-Object Hallucination in Vision Language Models
Xuwei Chen · Ziqiao Ma · Xuejun Zhang · Sihan Xu · Shengyi Qian · Jianing Yang · David Fouhey · Joyce Chai
I Don't Know: Explicit Modeling of Uncertainty with an [IDK] Token
Roi Cohen · Konstantin Dobler · Eden Biran · Gerard de Melo
Estimating the Hallucination Rate of Generative AI

Andrew Jesson · Nicolas Beltran Velez · Quentin Chu · Sweta Karlekar · Jannik Kossen · Yarin Gal · John Cunningham · David Blei
Truth is Universal: Robust Detection of Lies in LLMs
Lennart Bürger · Fred Hamprecht · Boaz Nadler
Alignment/AI Safety
Stress-Testing Capability Elicitation With Password-Locked Models
Ryan Greenblatt · Fabien Roger · Dmitrii Krashennikov · David Krueger
Gradient Cuff: Detecting Jailbreak Attacks on Large Language Models by Exploring Refusal Loss Landscapes
Xiaomeng Hu · Pin-Yu Chen · Tsung-Yi Ho
FLAME : Factuality-Aware Alignment for Large Language Models
Sheng-Chieh Lin · Luyu Gao · Barlas Oguz · Wenhan Xiong · Jimmy Lin · Scott Yih · Xilun Chen
Prejudice and Volatility: A Statistical Framework for Measuring Social Discrimination in Large Language Models
Yiran Liu · Ke Yang · Zehan Qi · Xiao Liu · Yang Yu · Cheng Xiang Zhai
Enhancing Multiple Dimensions of Trustworthiness in LLMs via Sparse Activation Control
Yuxin Xiao · Wan Chaoqun · Yonggang Zhang · Wenxiao Wang · Binbin Lin · Xiaofei He · Xu Shen · Jieping Ye
InfoRM: Mitigating Reward Hacking in RLHF via Information-Theoretic Reward Modeling
Yuchun Miao · Sen Zhang · Liang Ding · Rong Bao · Lefei Zhang · Dacheng Tao
Rule Based Rewards for Language Model Safety
Tong Mu · Alec Helyar · Johannes Heidecke · Joshua Achiam · Andrea Vallone · Ian Kivlichen · Molly Lin · Alex Beutel · John Schulman · Lilian Weng
Rethinking Misalignment in Vision-Language Model Adaptation from a Causal Perspective
Yanan Zhang · Jiangmeng Li · Lixiang Liu · Wenwen Qiang
CLAVE: An Adaptive Framework for Evaluating Values of LLM Generated Responses
Jing Yao · Xiaoyuan Yi · Xing Xie
Stratified Prediction-Powered Inference for Effective Hybrid Evaluation of Language Models
Adam Fisch · Joshua Maynez · R. Hofer · Bhuwan Dhingra · Amir Globerson · William Cohen
Refusal in Language Models Is Mediated by a Single Direction
Andy Arditi · Oscar Obeso · Aaquib Syed · Daniel Paleka · Nina Panickssery · Wes Gurnee · Neel Nanda
Stepwise Alignment for Constrained Language Model Policy Optimization
Akifumi Wachi · Thien Tran · Rei Sato · Takumi Tanabe · Youhei Akimoto
Watch Out for Your Agents! Investigating Backdoor Threats to LLM-Based Agents
Wenkai Yang · Xiaohan Bi · Yankai Lin · Sishuo Chen · Jie Zhou · Xu Sun
SpeechAlign: Aligning Speech Generation to Human Preferences
Dong Zhang · Zhaowei Li · Shimin Li · Xin Zhang · Pengyu Wang · Yaqian Zhou · Xipeng Qiu
Aligning to Thousands of Preferences via System Message Generalization
Seongyun Lee · Sue Hyun Park · Seungone Kim · Minjoon Seo
Vaccine: Perturbation-aware Alignment for Large Language Models against Harmful Fine-tuning Attack
Tiansheng Huang · Sihao Hu · Ling Liu
Privacy Backdoors: Enhancing Membership Inference through Poisoning Pre-trained Models
Yuxin Wen · Leo Marchyok · Sanghyun Hong · Jonas Geiping · Tom Goldstein · Nicholas Carlini
Improved Few-Shot Jailbreaking Can Circumvent Aligned Language Models and Their Defenses
Xiaosen Zheng · Tianyu Pang · Chao Du · Qian Liu · Jing Jiang · Min Lin
Shadowcast: Stealthy Data Poisoning Attacks Against Vision-Language Models
Yuancheng Xu · Jiarui Yao · Manli Shu · Yanchao Sun · Zichu Wu · Ning Yu · Tom Goldstein · Furong Huang
Large Language Model Unlearning via Embedding-Corrupted Prompts
Chris Liu · Yaxuan Wang · Jeffrey Flanigan · Yang Liu
Navigating the Safety Landscape: Measuring Risks in Finetuning Large Language Models
ShengYun Peng · Pin-Yu Chen · Matthew Hull · Duen Horng Chau
Rainbow Teaming: Open-Ended Generation of Diverse Adversarial Prompts
Mikayel Samvelyan · Sharath Chandra Raparthy · Andrei Lupu · Eric Hambro · Aram Markosyan · Manish Bhatt · Yuning Mao · Minqi Jiang · Jack Parker-Holder · Jakob Foerster · Tim Rocktäschel · Roberta Raileanu
Exploiting LLM Quantization
Kazuki Egashira · Mark Vero · Robin Staab · Jingxuan He · Martin Vechev
Selective Generation for Controllable Language Models
Minjae Lee · Kyungmin Kim · Taesoo Kim · Sangdon Park
PaCE: Parsimonious Concept Engineering for Large Language Models
Jinqi Luo · Tianjiao Ding · Kwan Ho Ryan Chan · Darshan Thaker · Aditya Chattopadhyay · Chris Callison-Burch · Rene Vidal
Evaluating Copyright Takedown Methods for Language Models
Boyi Wei · Weijia Shi · Yangsibo Huang · Noah Smith · Chiyuan Zhang · Luke Zettlemoyer · Kai Li · Peter Henderson
LT-Defense: Searching-free Backdoor Defense via Exploiting the Long-tailed Effect
Yixiao Xu · Binxing Fang · Mohan Li · Keke Tang · Zhihong Tian

INDICT: Code Generation with Internal Dialogues of Critiques for Both Security and Helpfulness
Hung Le · Doyen Sahoo · Yingbo Zhou · Caiming Xiong · Silvio Savarese
What Makes and Breaks Safety Fine-tuning? A Mechanistic Study
Samyak Jain · Ekdeep S Lubana · Kemal Oksuz · Tom Joy · Philip Torr · Amartya Sanyal · Puneet Dokania
Towards Safe Concept Transfer of Multi-Modal Diffusion via Causal Representation Editing
Peiran Dong · Bingjie WANG · Song Guo · Junxiao Wang · Jie ZHANG · Zicong Hong
Uncovering Safety Risks of Large Language Models through Concept Activation Vector
Zhihao Xu · Ruixuan HUANG · Changyu Chen · Xiting Wang
Prediction-Powered Ranking of Large Language Models
Ivi Chatzi · Eleni Straitouri · Suhas Thejaswi · Manuel Rodriguez
Decoding-Time Language Model Alignment with Multiple Objectives
Ruizhe Shi · Yifang Chen · Yushi Hu · Alisa Liu · Hanna Hajishirzi · Noah Smith · Simon Du
Fast Best-of-N Decoding via Speculative Rejection
Hanshi Sun · Momin Haider · Ruiqi Zhang · Huitao Yang · Jiahao Qiu · Ming Yin · Mengdi Wang · Peter Bartlett · Andrea Zanette
Semantic Density: Uncertainty Quantification for Large Language Models through Confidence Measurement in Semantic Space
Xin Qiu · Risto Miikkulainen
Grammar-Aligned Decoding
Kanghee Park · Jiayu Wang · Taylor Berg-Kirkpatrick · Nadia Polikarpova · Loris D'Antoni
TPR: Topology-Preserving Reservoirs for Generalized Zero-Shot Learning
Hui Chen · Yanbin Liu · Yongqiang Ma · Nanning Zheng · Xin Yu
The Art of Saying No: Contextual Noncompliance in Language Models
Faeze Brahman · Sachin Kumar · Vidhisha Balachandran · Pradeep Dasigi · Valentina Pyatkin · Abhilasha Ravichander · Sarah Wiegrefe · Nouha Dziri · Khyathi Chandu · Jack Hessel · Yulia Tsvetkov · Noah Smith · Yejin Choi · Hannaneh Hajishirzi
Mission Impossible: A Statistical Perspective on Jailbreaking LLMs
Jingtong Su · Julia Kempe · Karen Ullrich
Secret Collusion among AI Agents: Multi-Agent Deception via Steganography
Sumeet Motwani · Mikhail Baranchuk · Martin Strohmeier · Vijay Bolina · Philip Torr · Lewis Hammond · Christian Schroeder de Witt
Protecting Your LLMs with Information Bottleneck
Zichuan Liu · Zefan Wang · Linjie Xu · Jinyu Wang · Lei Song · Tianchun Wang · Chunlin Chen · Wei Cheng · Jiang Bian
Spectral Editing of Activations for Large Language Model Alignment
Yifu QIU · Zheng Zhao · Yftah Ziser · Anna Korhonen · Edoardo Maria Ponti · Shay Cohen
Apathetic or Empathetic? Evaluating LLMs' Emotional Alignments with Humans
Jen-Tse Huang · Man Ho LAM · Eric John Li · Shujie Ren · Wenxuan Wang · Wenxiang Jiao · Zhaopeng Tu · Michael R Lyu
PrivacyLens: Evaluating Privacy Norm Awareness of Language Models in Action
Yijia Shao · Tianshi Li · Weiyan Shi · Yanchen Liu · Diyi Yang
LLM Evaluators Recognize and Favor Their Own Generations
Arjun Panickssery · Samuel Bowman · Shi Feng
Aligning Large Language Models with Representation Editing: A Control Perspective
Lingkai Kong · Haorui Wang · Wenhao Mu · Yuanqi Du · Yuchen Zhuang · Yifei Zhou · Yue Song · Rongzhi Zhang · Kai Wang · Chao Zhang
DAGER: Exact Gradient Inversion for Large Language Models
Ivo Petrov · Dimitar I. Dimitrov · Maximilian Baader · Mark Müller · Martin Vechev
Large Language Model Unlearning
Yuanshun Yao · Xiaojun Xu · Yang Liu
BackdoorAlign: Mitigating Fine-tuning based Jailbreak Attack with Backdoor Enhanced Safety Alignment
Jiong Xiao Wang · Jiazhao LI · Yiquan Li · Xiangyu Qi · Junjie Hu · Sharon Li · Patrick McDaniel · Muhao Chen · Bo Li · Chaowei Xiao
Tree of Attacks: Jailbreaking Black-Box LLMs Automatically
Anay Mehrotra · Manolis Zampetakis · Paul Kassianik · Blaine Nelson · Hyrum Anderson · Yaron Singer · Amin Karbasi
Efficient Adversarial Training in LLMs with Continuous Attacks
Louis-Pascal Xhonneux · Alessandro Sordani · Stephan Günnemann · Gauthier Gidel · Leo Schwinn
Can Models Learn Skill Composition from Examples?
Haoyu Zhao · Simran Kaur · Dingli Yu · Anirudh Goyal · Sanjeev Arora
ART: Automatic Red-teaming for Text-to-Image Models to Protect Benign Users
Guanlin Li · Kangjie Chen · Shudong Zhang · Jie Zhang · Tianwei Zhang
Panacea: Pareto Alignment via Preference Adaptation for LLMs
Yifan Zhong · Chengdong Ma · Xiaoyuan Zhang · Ziran Yang · Haojun Chen · Qingfu Zhang · Siyuan Qi · Yaodong Yang
ReMoDetect: Reward Models Recognize Aligned LLM's Generations
Hyunseok Lee · Jihoon Tack · Jinwoo Shin
WAGLE: Strategic Weight Attribution for Effective and Modular Unlearning in Large Language Models

Jinghan Jia · Jiancheng Liu · Yihua Zhang · Parikshit Ram · Nathalie Baracaldo · Sijia Liu
Representation Noising: A Defence Mechanism Against Harmful Finetuning
Domenic Rosati · Jan Wehner · Kai Williams · Lukasz Bartoszcze · Robie Gonzales · carsten maple · Subhabrata Majumdar · Hassan Sajjad · Frank Rudzicz
Unveiling the Bias Impact on Symmetric Moral Consistency of Large Language Models
Ziyi Zhou · Xinwei Guo · Jiashi Gao · Xiangyu Zhao · Shiyao Zhang · Xin Yao · Xuetao Wei
Open LLMs are Necessary for Current Private Adaptations and Outperform their Closed Alternatives
Vincent Hanke · Tom Blanchard · Franziska Boenisch · Iyiola Olatunji · Michael Backes · Adam Dziedzić
When is Multicalibration Post-Processing Necessary?
Dutch Hansen · Siddhartha Devic · Preetum Nakkiran · Vatsal Sharan
BoNBoN Alignment for Large Language Models and the Sweetness of Best-of-n Sampling
Lin Gui · Cristina Garbacea · Victor Veitch
MAGNET: Improving the Multilingual Fairness of Language Models with Adaptive Gradient-Based Tokenization
Orevaoghene Ahia · Sachin Kumar · Hila Gonen · Valentin Hofmann · Tomasz Limisiewicz · Yulia Tsvetkov · Noah Smith
Fight Back Against Jailbreaking via Prompt Adversarial Tuning
Yichuan Mo · Yuji Wang · Zeming Wei · Yisen Wang
Efficient LLM Jailbreak via Adaptive Dense-to-sparse Constrained Optimization
Kai Hu · Weichen Yu · Yining Li · Tianjun Yao · Xiang Li · Wenhe Liu · Lijun Yu · Zhiqiang Shen · Kai Chen · Matt Fredrikson
Alignment for Honesty
Yuqing Yang · Ethan Chern · Xipeng Qiu · Graham Neubig · Pengfei Liu
Robust Prompt Optimization for Defending Language Models Against Jailbreaking Attacks
Andy Zhou · Bo Li · Haohan Wang
Pandora's Box: Towards Building Universal Attackers against Real-World Large Vision-Language Models
Daizong Liu · Mingyu Yang · Xiaoye Qu · Pan Zhou · Xiang Fang · Keke Tang · Yao Wan · Lichao Sun
MoGU: A Framework for Enhancing Safety of LLMs While Preserving Their Usability
YANRUI DU · Sendong Zhao · Danyang Zhao · Ming Ma · Yuhan Chen · Liangyu Huo · Qing Yang · Dongliang Xu · Bing Qin
ALI-Agent: Assessing LLMs' Alignment with Human Values via Agent-based Evaluation
jingnan zheng · Han Wang · An Zhang · Nguyen Duy Tai · Jun Sun · Tat-Seng Chua
HonestLLM: Toward an Honest and Helpful Large Language Model
Gao Chujie · Siyuan Wu · Yue Huang · Dongping Chen · Qihui Zhang · Zhengyan Fu · Yao Wan · Lichao Sun · Xiangliang Zhang
A Theoretical Understanding of Self-Correction through In-context Alignment
Yifei Wang · Yuyang Wu · Zeming Wei · Stefanie Jegelka · Yisen Wang
Bias Amplification in Language Model Evolution: An Iterated Learning Perspective
Yi Ren · Shangmin Guo · Linlu Qiu · Bailin Wang · Danica J. Sutherland
Private Attribute Inference from Images with Vision-Language Models
Batuhan Tömekçe · Mark Vero · Robin Staab · Martin Vechev
Introspective Planning: Aligning Robots' Uncertainty with Inherent Task Ambiguity
Kaiqu Liang · Zixu Zhang · Jaime Fisac
Noise Contrastive Alignment of Language Models with Explicit Rewards
Huayu Chen · Guande He · Lifan Yuan · Ganqu Cui · Hang Su · Jun Zhu
ProgressGym: Alignment with a Millennium of Moral Progress
Tianyi (Alex) Qiu · Yang Zhang · Xuchuan Huang · Jasmine Li · Jiaming Ji · Yaodong Yang
GuardT2I: Defending Text-to-Image Models from Adversarial Prompts
Yijun Yang · Ruiyuan Gao · Xiao Yang · Jianyuan Zhong · Qiang Xu
UniBias: Unveiling and Mitigating LLM Bias through Internal Attention and FFN Manipulation
Hanzhang Zhou · Zijian Feng · Zixiao Zhu · Junlang Qian · Kezhi Mao
Incentivizing Quality Text Generation via Statistical Contracts
Eden Saig · Ohad Einav · Inbal Talgam-Cohen
Unelicitable Backdoors via Cryptographic Transformer Circuits
Andis Draguns · Andrew Gritsevskiy · Sumeet Motwani · Christian Schroeder de Witt
PrivAuditor: Benchmarking Data Protection Vulnerabilities in LLM Adaptation Techniques
Derui Zhu · Dingfan Chen · Xiongfei Wu · Jiahui Geng · Zhuo Li · Jens Grossklags · Lei Ma
Query-Based Adversarial Prompt Generation
Jonathan Hayase · Ema Borevković · Nicholas Carlini · Florian Tramer · Milad Nasr
Improving Alignment and Robustness with Circuit Breakers
Andy Zou · Long Phan · Justin Wang · Derek Duenas · Maxwell Lin · Maksym Andriushchenko · J. Zico Kolter · Matt Fredrikson · Dan Hendrycks
Perplexity-aware Correction for Robust Alignment with Noisy Preferences
Keyi Kong · Xilie Xu · Di Wang · Jingfeng ZHANG · Mohan Kankanhalli
Injecting Undetectable Backdoors in Obfuscated Neural Networks and Language Models
Alkis Kalavasis · Amin Karbasi · Argyris Oikonomou · Katerina Sotiraki · Grigoris Veleghas · Manolis Zampetakis

Weak-to-Strong Search: Align Large Language Models via Searching over Small Language Models
Zhanhui Zhou · Zhixuan Liu · Jie Liu · Zhichen Dong · Chao Yang · Yu Qiao
MetaAligner: Towards Generalizable Multi-Objective Alignment of Language Models
Kailai Yang · Zhiwei Liu · Qianqian Xie · Jimin Huang · Tianlin Zhang · Sophia Ananiadou
Aligner: Efficient Alignment by Learning to Correct
Jiaming Ji · Boyuan Chen · Hantao Lou · Donghai Hong · Borong Zhang · Xuehai Pan · Tianyi (Alex) Qiu · Juntao Dai · Yaodong Yang
Lisa: Lazy Safety Alignment for Large Language Models against Harmful Fine-tuning Attack
Tiansheng Huang · Sihao Hu · Fatih Ilhan · Selim Tekin · Ling Liu
Dataset and Lessons Learned from the 2024 SaTML LLM Capture-the-Flag Competition
Edoardo Debenedetti · Javier Rando · Daniel Paleka · Silaghi Florin · Dragos Albastroiu · Niv Cohen · Yuval Lemberg · Reshmi Ghosh · Rui Wen · Ahmed Salem · Giovanni Cherubin · Santiago Zanella-Beguelin · Robin Schmid · Victor Klemm · Takahiro Miki · Chenhao Li · Stefan Kraft · Mario Fritz · Florian Tramer · Sahar Abdelnabi · Lea Schönherr
SelfCodeAlign: Self-Alignment for Code Generation
Yuxiang Wei · Federico Cassano · Jiawei Liu · Yifeng Ding · Naman Jain · Zachary Mueller · Harm de Vries · Leandro Von Werra · Arjun Guha · LINGMING ZHANG
Jailbreaking Large Language Models Against Moderation Guardrails via Cipher Characters
Haibo Jin · Andy Zhou · Joe Menke · Haohan Wang
Provably Mitigating Overoptimization in RLHF: Your SFT Loss is Implicitly an Adversarial Regularizer
Zhihan Liu · Miao Lu · Shenao Zhang · Boyi Liu · Hongyi Guo · Yingxiang Yang · Jose Blanchet · Zhaoran Wang
EAI: Emotional Decision-Making of LLMs in Strategic Games and Ethical Dilemmas
Mikhail Mozikov · Nikita Severin · Valeria Bodishtianu · Maria Glushanina · Ivan Nasonov · Daniil Orekhov · Pekhotin Vladislav · Ivan Makovetskiy · Mikhail Baklashkin · Vasily Lavrentyev · Akim Tsvigun · Denis Turdakov · Tatiana Shavrina · Andrey Savchenko · Ilya Makarov
One-Shot Safety Alignment for Large Language Models via Optimal Dualization
Xinmeng Huang · Shuo Li · Edgar Dobriban · Osbert Bastani · Hamed Hassani · Dongsheng Ding
Alignment at Pre-training! Towards Native Alignment for Arabic LLMs
Juhao Liang · Zhenyang Cai · Jianqing Zhu · Huang Huang · Kewei Zong · Bang An · Mosen Alharthi · Juncal He · Lian Zhang · Haizhou Li · Benyou Wang · Jinchao Xu
Conformal Alignment: Knowing When to Trust Foundation Models with Guarantees
Yu Gui · Ying Jin · Zhimei Ren
Cal-DPO: Calibrated Direct Preference Optimization for Language Model Alignment
Teng Xiao · Yige Yuan · Huaisheng Zhu · Mingxiao Li · Vasant Honavar
Learning Human-like Representations to Enable Learning Human Values
Andrea Wynn · Ilia Sucholutsky · Tom Griffiths
Benchmark and Dataset
DACO: Towards Application-Driven and Comprehensive Data Analysis via Code Generation
Xueqing Wu · Rui Zheng · Jingzhen Sha · Te-Lin Wu · Hanyu Zhou · Tang Mohan · Kai-Wei Chang · Nanyun Peng · Haoran Huang
UltraMedical: Building Specialized Generalists in Biomedicine
Kaiyan Zhang · Sihang Zeng · Ermo Hua · Ning Ding · Zhang-Ren Chen · Zhiyuan Ma · Haoxin Li · Ganqu Cui · Biqing Qi · Xuekai Zhu · Xingtai Lv · Hu Jinfang · Zhiyuan Liu · Bowen Zhou
WizardArena: Post-training Large Language Models via Simulated Offline Chatbot Arena
HAIPENG LUO · Qingfeng Sun · Can Xu · Pu Zhao · Qingwei Lin · Jian-Guang Lou · Shifeng Chen · Yansong Tang · Weizhu Chen
Generating Code World Models with Large Language Models Guided by Monte Carlo Tree Search
Nicola Dainese · Matteo Merler · Minttu Alakuijala · Pekka Marttinen
UDA: A Benchmark Suite for Retrieval Augmented Generation in Real-World Document Analysis
Yulong Hui · YAO LU · Huanchen Zhang
ComBack: A Versatile Dataset for Enhancing Compiler Backend Development Efficiency
Ming Zhong · FANG LYU · Lulin Wang · Hongna Geng · Lei Qiu · Huimin Cui · Xiaobing Feng
Empowering and Assessing the Utility of Large Language Models in Crop Science
Hang Zhang · Jiawei SUN · Renqi Chen · Wei Liu · Zhonghang Yuan · Xinzhe Zheng · Zhefan Wang · Zhiyuan Yang · Hang Yan · Han-Sen Zhong · Xiqing Wang · Fan Yang · Nanqing Dong · Wanli Ouyang
SocraticLM: Exploring Socratic Personalized Teaching with Large Language Models
Jiayu Liu · Zhenya Huang · Tong Xiao · Jing Sha · Jinze Wu · Qi Liu · Shijin Wang · Enhong Chen
DTGB: A Comprehensive Benchmark for Dynamic Text-Attributed Graphs
Jiasheng Zhang · Jialin Chen · Menglin Yang · Aosong Feng · Shuang Liang · Jie Shao · Rex Ying
Automating Dataset Updates Towards Reliable and Timely Evaluation of Large Language Models
Jiahao Ying · Yixin Cao · Yushi Bai · QIANRU SUN · Bo Wang · Wei Tang · Zhaojun Ding · Yizhe Yang · Xuanjing Huang · Shuicheng Yan
Game-Traversal-Benchmark: Evaluating Planning Abilities Of Large Language Models Via Traversing 2D Game Maps

Muhammad Umair Nasir · Steven James · Julian Togelius
The PRISM Alignment Dataset: What Participatory, Representative and Individualised Human Feedback Reveals About the Subjective and Multicultural Alignment of Large Language Models
Hannah Rose Kirk · Alexander Whitefield · Paul Rottger · Andrew M. Bean · Katerina Margatina · Rafael Mosquera-Gomez · Juan Ciro · Max Bartolo · Adina Williams · He He · Bertie Vidgen · Scott Hale
MARVEL: Multidimensional Abstraction and Reasoning through Visual Evaluation and Learning
Yifan Jiang · jiarui zhang · Kexuan Sun · Zhivar Sourati Hassan Zadeh · Kian Ahrabian · Kaixin Ma · Filip Ilievski · Jay Pujara
HW-GPT-Bench: Hardware-Aware Architecture Benchmark for Language Models
Rhea Sukthanker · Arber Zela · Benedikt Staffler · Aaron Klein · Lennart Purucker · Jörg Franke · Frank Hutter
FinBen: An Holistic Financial Benchmark for Large Language Models
Qianqian Xie · Weiguang Han · Zhengyu Chen · Ruoyu Xiang · Xiao Zhang · Yueru He · Mengxi Xiao · Dong Li · Yongfu Dai · Duanyu Feng · Yijing Xu · Haoqiang Kang · Ziyang Kuang · Chenhan Yuan · Kailai Yang · Zheheng Luo · Tianlin Zhang · Zhiwei Liu · GUOJUN XIONG · Zhiyang Deng · Yuechen Jiang · Zhiyuan Yao · Haohang Li · Yangyang Yu · Gang Hu · Huang Jiajia · Xiaoyang Liu · Alejandro Lopez-Lira · Benyou Wang · Yanzhao Lai · Hao Wang · Min Peng · Sophia Ananiadou · Jimin Huang
DivSafe: Evaluating the Generalization of LLM Safety Training Across Diverse Tasks and Prompt Types
Yutao Mou · Shikun Zhang · Wei Ye
DetectEval: Benchmarking LLM-Generated Text Detection in Real-World Scenarios
Junchao Wu · Runzhe Zhan · Derek Wong · Shu Yang · Xinyi Yang · Yulin Yuan · Lidia Chao
Evaluating Large Vision-and-Language Models on Children's Mathematical Olympiads
Anoop Cherian · Kuan-Chuan Peng · Suhas Lohit · Joanna Matthiesen · Kevin Smith · Josh Tenenbaum
Can LLMs Solve Molecule Puzzles? A Multimodal Benchmark for Molecular Structure Elucidation
Kehan Guo · Bozhao Nan · Yujun Zhou · Taicheng Guo · Zhichun Guo · Mihir Surve · Zhenwen Liang · Nitesh Chawla · Olaf Wiest · Xiangliang Zhang
Stress-Testing Long-Context Language Models with Lifelong ICL and Task Haystack
Xiaoyue Xu · Qinyuan Ye · Xiang Ren
RedPajama: an Open Dataset for Training Large Language Models
Maurice Weber · Dan Fu · Quentin Anthony · Yonatan Oren · Shane Adams · Anton Alexandrov · Xiaozhong Lyu · Huu Nguyen · Xiaozhe Yao · Virginia Adams · Ben Athiwaratkun · Rahul Chalamala · Kezhen Chen · Max Ryabinin · Tri Dao · Percy Liang · Christopher Ré · Irina Rish · Ce Zhang
A Sober Look at the Robustness of CLIPs to Spurious Features
Qizhou Wang · Yong Lin · Yongqiang Chen · Ludwig Schmidt · Bo Han · Tong Zhang
SpatialRGPT: Grounded Spatial Reasoning in Vision-Language Models
An-Chieh Cheng · Hongxu Yin · Yang Fu · Qiushan Guo · Ruihan Yang · Jan Kautz · Xiaolong Wang · Sifei Liu
VERIFIED: A Video Corpus Moment Retrieval Benchmark for Fine-Grained Video Understanding
Houlun Chen · Xin Wang · Hong Chen · Zeyang Zhang · Wei Feng · Bin Huang · Jia Jia · Wenwu Zhu
SciInstruct: a Self-Reflective Instruction Annotated Dataset for Training Scientific Language Models
Dan Zhang · Ziniu Hu · Sining Zhoubian · Zhengxiao Du · Kaiyu Yang · Zihan Wang · Yisong Yue · Yuxiao Dong · Jie Tang
Visual CoT: Advancing Multi-Modal Language Models with a Comprehensive Dataset and Benchmark for Chain-of-Thought Reasoning
Hao Shao · Shengju Qian · Han Xiao · Guanglu Song · ZHUOFAN ZONG · Letian Wang · Yu Liu · Hongsheng Li
SWT-Bench: Testing and Validating Real-World Bug-Fixes with Code Agents
Niels Mündler · Mark Müller · Jingxuan He · Martin Vechev
SafeWorld: Geo-Diverse Safety Alignment
Da Yin · Haoyi Qiu · Kung-Hsiang Huang · Kai-Wei Chang · Nanyun Peng
A Careful Examination of Large Language Model Performance on Grade School Arithmetic
Hugh Zhang · Jeff Da · Dean Lee · Vaughn Robinson · Catherine Wu · William Song · Tiffany Zhao · Pranav Raja · Charlotte Zhuang · Dylan Slack · Qin Lyu · Sean Hendryx · Russell Kaplan · Michele Lunati · Summer Yue
Hints-In-Browser: Benchmarking Language Models for Programming Feedback Generation
Nachiket Kotalwar · Alkis Gotovos · Adish Singla
NaturalBench: Evaluating Vision-Language Models on Natural Adversarial Samples
Baiqi Li · Zhiqiu Lin · WENXUAN PENG · Jean de Dieu Nyandwi · Daniel Jiang · Zixian Ma · Simran Khanuja · Ranjay Krishna · Graham Neubig · Deva Ramanan
Humor in AI: Massive Scale Crowd-Sourced Preferences and Benchmarks for Cartoon Captioning
Jifan Zhang · Lalit Jain · Yang Guo · Jiayi Chen · Kuan Zhou · Siddharth Suresh · Andrew Wagenmaker · Scott Sievert · Timothy T Rogers · Kevin Jamieson · Bob Mankoff · Robert Nowak
DiReCT: Diagnostic Reasoning for Clinical Notes via Large Language Models
Bowen Wang · Jiuyang Chang · Yiming Qian · Guoxin Chen · Junhao Chen · Zhouqiang Jiang · Jiahao Zhang · Yuta Nakashima · Hajime Nagahara
Decision-Making Behavior Evaluation Framework for LLMs under Uncertain Context
Jingru (Jessica) Jia · Zehua Yuan · Junhao Pan · Paul McNamara · Deming Chen
Measuring Dejavu Memorization Efficiently

Narine Kokhlikyan · Bargav Jayaraman · Florian Bordes · Chuan Guo · Kamalika Chaudhuri
DART-Eval: A Comprehensive DNA Language Model Evaluation Benchmark on Regulatory DNA
Aman Patel · Arpita Singhal · Austin Wang · Anusri Pampari · Maya Kasowski · Anshul Kundaje
Multi-language Diversity Benefits Autoformalization
Albert Q. Jiang · Wenda Li · Mateja Jamnik
Elo Uncovered: Robustness and Best Practices in Language Model Evaluation
Meriem Boubdir · Edward Kim · Beyza Ermis · Sara Hooker · Marzieh Fadaee
Embodied Agent Interface: Benchmarking LLMs for Embodied Decision Making
Manling Li · Shiyu Zhao · Qineng Wang · Kangrui Wang · Yu Zhou · Sanjana Srivastava · Cem Gokmen · Tony Lee · Erran Li Li · Ruohan Zhang · Weiyu Liu · Percy Liang · Fei-Fei Li · Jiajun Wu · Jiayuan Mao
AsEP: Benchmarking Deep Learning Methods for Antibody-specific Epitope Prediction
ChunNan Liu · Lilian Denzler · Yihong Chen · Andrew Martin · Brooks Paige
StreamBench: Towards Benchmarking Continuous Improvement of Language Agents
Cheng-Kuang Wu · Zhi Rui Tam · Chieh-Yen Lin · Yun-Nung (Vivian) Chen · Hung-yi Lee
OlympicArena: Benchmarking Multi-discipline Cognitive Reasoning for Superintelligent AI
Zhen Huang · Zengzhi Wang · Shijie Xia · Xuefeng Li · Haoyang Zou · Ruijie Xu · Run-Ze Fan · Lyumanshan Ye · Ethan Chern · Yixin Ye · Yikai Zhang · Yuqing Yang · Ting Wu · Binjie Wang · Shichao Sun · Yang Xiao · Yiyuan Li · Fan Zhou · Steffi Chern · Yiwei Qin · Yan Ma · Jiadi Su · Yixiu Liu · Yuxiang Zheng · Shaoting Zhang · Dahua Lin · Yu Qiao · Pengfei Liu
SpreadsheetBench: Towards Challenging Real World Spreadsheet Manipulation
Zeyao Ma · Bohan Zhang · Jing Zhang · Jifan Yu · Xiaokang Zhang · Xiaohan Zhang · Sijia Luo · Xi Wang · Jie Tang
MLLMGuard: A Multi-dimensional Safety Evaluation Suite for Multimodal Large Language Models
Tianle Gu · Zeyang Zhou · Kexin Huang · Liang Dandan · Yixu Wang · Haiquan Zhao · Yuanqi Yao · xingge qiao · Keqing wang · Yujiu Yang · Yan Teng · Yu Qiao · Yingchun Wang
MMLONGBENCH-DOC: Benchmarking Long-context Document Understanding with Visualizations
Yubo Ma · Yuhang Zang · Liangyu Chen · Meiqi Chen · Yizhu Jiao · Xinze Li · Xinyuan Lu · Ziyu Liu · Yan Ma · Xiaoyi Dong · Pan Zhang · Liangming Pan · Yu-Gang Jiang · Jiaqi Wang · Yixin Cao · Aixin Sun
MMLU-Pro: A More Robust and Challenging Multi-Task Language Understanding Benchmark
Yubo Wang · Xueguang Ma · Ge Zhang · Yuansheng Ni · Abhranil Chandra · Shiguang Guo · Weiming Ren · Aaran Arulraj · Xuan He · Ziyang Jiang · Tianle Li · Max KU · Kai Wang · Alex Zhuang · Rongqi Fan · Xiang Yue · Wenhu Chen
WorkArena++: Towards Compositional Planning and Reasoning-based Common Knowledge Work Tasks
Léo Boisvert · Megh Thakkar · Maxime Gasse · Massimo Caccia · Thibault de Chezelles · Quentin Cappart · Nicolas Chapados · Alexandre Lacoste · Alexandre Drouin
BeanCounter: A low-toxicity, large-scale, and open dataset of business-oriented text
Siyan Wang · Bradford Levy
Cross-Care: Assessing the Healthcare Implications of Pre-training Data on Language Model Bias
Shan Chen · Jack Gallifant · Mingye Gao · Pedro Moreira · Nikolaj Munch · Ajay Muthukkumar · Arvind Rajan · Jaya Kolluri · Amelia Fiske · Janna Hastings · Hugo Aerts · Brian Anthony · Leo Anthony Celi · William La Cava · Danielle Bitterman
SubjEctive-QA: A dataset for the subjective evaluation of answers in Earnings Call Transcripts (ECTs)
Huzaifa Pardawala · Siddhant Sukhani · Veer Kejriwal · Rohan Bhasin · Abhishek Pillai · Dhruv Adha · Tarun Mandapati · Andrew DiBiasio · Agam Shah · Sudheer Chava
BEACON: Benchmark for Comprehensive RNA Tasks and Language Models
Yuchen Ren · Zhiyuan Chen · Lifeng Qiao · Hongtai Jing · Yuchen Cai · Sheng Xu · Peng Ye · Xinzhu Ma · Siqi Sun · Hongliang Yan · Dong Yuan · Wanli Ouyang · Xihui Liu
EHRCon: Dataset for Checking Consistency between Unstructured Notes and Structured Tables in Electronic Health Records
Yeonsu Kwon · Jiho Kim · Gyubok Lee · Seongsu Bae · Daeun Kyung · Wonchul Cha · Tom Pollard · Alistair Johnson · Edward Choi
ConvBench: A Multi-Turn Conversation Evaluation Benchmark with Hierarchical Ablation Capability for Large Vision-Language Models
Shuo Liu · Kaining Ying · Hao Zhang · yue yang · Yuqi Lin · Tianle Zhang · Chuanhao Li · Yu Qiao · Ping Luo · Wenqi Shao · Kaipeng Zhang
Diff-eRank: A Novel Rank-Based Metric for Evaluating Large Language Models
Lai Wei · Zhiquan Tan · Chenghai Li · Jindong Wang · Weiran Huang
BLEND: A Benchmark for LLMs on Everyday Knowledge in Diverse Cultures and Languages
Junho Myung · Nayeon Lee · Yi Zhou · Jiho Jin · Rifki Putri · Dimosthenis Antypas · Hsuvas Borkakoty · Eunsu Kim · Carla Perez-Almendros · Abinew Ali Ayele · Victor Gutierrez Basulto · Yazmin Ibanez-Garcia · Hwaran Lee · Shamsuddeen H Muhammad · Kiwoong Park · Anar Rzayev · Nina White · Seid Muhie Yimam · Mohammad Taher Pilehvar · Nedjma Ousidhoum · Jose Camacho-Collados · Alice Oh
AgentBoard: An Analytical Evaluation Board of Multi-turn LLM Agents
Ma Chang · Junlei Zhang · Zhihao Zhu · Cheng Yang · Yujiu Yang · Yaohui Jin · Zhenzhong Lan · Lingpeng Kong · Junxian He
Large Language Models Play StarCraft II: Benchmarks and A Chain of Summarization Approach
Weiyu Ma · Qirui Mi · Yongcheng Zeng · Xue Yan · Runji Lin · Yuqiao Wu · Jun Wang · Haifeng Zhang

MR-Ben: A Meta-Reasoning Benchmark for Evaluating System-2 Thinking in LLMs
Zhongshen Zeng · Yinhong Liu · Yingjia Wan · Jingyao Li · Pengguang Chen · Jianbo Dai · Yuxuan Yao · Rongwu Xu · Zehan Qi · Wanru Zhao · Linling Shen · Jianqiao Lu · Haochen Tan · Yukang Chen · Hao Zhang · Zhan Shi · Bailin Wang · Zhijiang Guo · Jiaya Jia
Multivariate Stochastic Dominance via Optimal Transport and Applications to Models Benchmarking
Gabriel Rioux · Apoorva Nitsure · Mattia Rigotti · Kristjan Greenewald · Youssef Mroueh
IDGen: Item Discrimination Induced Prompt Generation for LLM Evaluation
Fan Lin · Shuyi Xie · Yong Dai · Wenlin Yao · TianJiao Lang · Yu Zhang
OpenDebateEvidence: A Massive-Scale Argument Mining and Summarization Dataset
Allen Roush · Yusuf Shabazz · Arvind Balaji · Peter Zhang · Stefano Mezza · Markus Zhang · Sanjay Basu · Sriram Vishwanath · Ravid Shwartz-Ziv
MedSafetyBench: Evaluating and Improving the Medical Safety of Large Language Models
Tessa Han · Aounon Kumar · Chirag Agarwal · Himabindu Lakkaraju
SciCode: A Research Coding Benchmark Curated by Scientists
Minyang Tian · Luyu Gao · Shizhuo Zhang · Xinan Chen · Cunwei Fan · Xuefei Guo · Roland Haas · Pan Ji · Kittithat Krongchon · Yao Li · Shengyan Liu · Di Luo · Yutao Ma · HAO TONG · Kha Trinh · Chenyu Tian · Zihan Wang · Bohao Wu · Shengzhu Yin · Minhui Zhu · Kilian Lieret · Yanxin Lu · Genglin Liu · Yufeng Du · Tianhua Tao · Ofir Press · Jamie Callan · Eliu Huerta · Hao Peng
ReXTime: A Benchmark Suite for Reasoning-Across-Time in Videos
JR-JEN CHEN · Yu-Chien Liao · Hsi-Che Lin · Yu-Chu Yu · Yen-Chun Chen · Frank Wang
Unveiling the Tapestry of Consistency in Large Vision-Language Models
Yuan Zhang · Fei xiao · Tao Huang · Chun-Kai Fan · Hongyuan Dong · Jiawen Li · Jiacong Wang · Kuan Cheng · Shanghang Zhang · Haoyuan Guo
MixEval: Deriving Wisdom of the Crowd from LLM Benchmark Mixtures
Jinjie Ni · Fuzhao Xue · Xiang Yue · Yuntian Deng · Mahir Shah · Kabir Jain · Graham Neubig · Yang You
A Synthetic Dataset for Personal Attribute Inference
Hanna Yukhymenko · Robin Staab · Mark Vero · Martin Vechev
Visual Riddles: a Commonsense and World Knowledge Challenge for Large Vision and Language Models
Nitzan Bitton Guetta · Aviv Slobodkin · Aviya Maimon · Eliya Habba · Royi Rassin · Yonatan Bitton · Idan Szpektor · Amir Globerson · Yuval Elovici
When LLMs Meet Cunning Texts: A Fallacy Understanding Benchmark for Large Language Models
Yinghui Li · Qingyu Zhou · Yuanzhen Luo · Shirong Ma · Yangning Li · Hai-Tao Zheng · Xuming Hu · Philip S Yu
ImageNet3D: Towards General-Purpose Object-Level 3D Understanding
Wufei Ma · Guanning Zeng · Guofeng Zhang · Qihao Liu · Letian Zhang · Adam Kortylewski · Yaoyao Liu · Alan Yuille
Can Large Language Models Analyze Graphs like Professionals? A Benchmark and Dataset
Xin Li · Weize Chen · Qizhi Chu · Haopeng Li · Zhaojun Sun · Ran Li · Chen Qian · Yiwei Wei · Chuan Shi · Zhiyuan Liu · Maosong Sun · Cheng Yang
WikiContradict: A Benchmark for Evaluating LLMs on Real-World Knowledge Conflicts from Wikipedia
Yufang Hou · Alessandra Pascale · Javier Carnerero-Cano · Tigran Tchrakian · Radu Marinescu · Elizabeth Daly · Inkit Padhi · Prasanna Sattigeri
WildVision: Evaluating Vision-Language Models in the Wild with Human Preferences
Yujie Lu · Dongfu Jiang · Wenhui Chen · William Yang Wang · Yejin Choi · Bill Yuchen Lin
Mercury: A Code Efficiency Benchmark for Code Large Language Models
Mingzhe Du · Anh Tuan Luu · Bin Ji · Qian Liu · See-Kiong Ng
VHELM: A Holistic Evaluation of Vision Language Models
Tony Lee · Haoqin Tu · Chi Heem Wong · Wenhao Zheng · Yiyang Zhou · Yifan Mai · Josselin Roberts · Michihiro Yasunaga · Huaxiu Yao · Cihang Xie · Percy Liang
μBench: A Microscopy Benchmark for Vision-Language Understanding
Alejandro Lozano · Jeffrey Nirschl · James Burgess · Sanket Rajan Gupte · Yuhui Zhang · Alyssa Unell · Serena Yeung
TaskBench: Benchmarking Large Language Models for Task Automation
Yongliang Shen · Kaitao Song · Xu Tan · Wenqi Zhang · Kan Ren · Siyu Yuan · Weiming Lu · Dongsheng Li · Yueting Zhuang
Right this way: Can VLMs Guide Us to See More to Answer Questions?
Li Liu · Diji Yang · Sijia Zhong · Kalyana Suma Sree Tholeti · Lei Ding · Yi Zhang · Leilani Gilpin
Cracking the Code of Juxtaposition: Can AI Models Understand the Humorous Contradictions
Zhe Hu · Tuo Liang · Jing Li · Yiren Lu · Yunlai Zhou · Yiran Qiao · Jing Ma · Yu Yin
Bag of Tricks: Benchmarking of Jailbreak Attacks on LLMs
Zhao Xu · Fan LIU · Hao Liu
CTIBench: A Benchmark for Evaluating LLMs in Cyber Threat Intelligence
Md Tanvirul Alam · Dipkamal Bhusal · Le Nguyen · Nidhi Rastogi
HelpSteer 2: Open-source dataset for training top-performing reward models
Zhilin Wang · Yi Dong · Olivier Delalleau · Jiaqi Zeng · Gerald Shen · Daniel Egert · Jimmy Zhang · Makesh Narsimhan Sreedhar · Oleksii Kuchaiev

OpenMathInstruct-1: A 1.8 Million Math Instruction Tuning Dataset
Shubham Toshniwal · Ivan Moshkov · Sean Narenthiran · Daria Gitman · Fei Jia · Igor Gitman
EHRNoteQA: An LLM Benchmark for Real-World Clinical Practice Using Discharge Summaries
Sunjun Kweon · Jiyoun Kim · Heeyoung Kwak · Dongchul Cha · Hangyul Yoon · Kwang Kim · Jeewon Yang · Seunghyun Won · Edward Choi
PertEval: Unveiling Real Knowledge Capacity of LLMs with Knowledge-Invariant Perturbations
Jiatong Li · Renjun Hu · Kunzhe Huang · Yan Zhuang · Qi Liu · Mengxiao Zhu · Xing Shi
kGym: A Platform and Dataset to Benchmark Large Language Models on Linux Kernel Crash Resolution
Alex Mathai · Chenxi Huang · Petros Maniatis · Aleksandr Nogikh · Franjo Ivančić · Junfeng Yang · Baishakhi Ray
VisMin: Visual Minimal-Change Understanding
Rabiul Awal · Saba Ahmadi · LE ZHANG · Aishwarya Agrawal
Benchmarking Complex Instruction-Following with Multiple Constraints Composition
Bosi Wen · Pei Ke · Xiaotao Gu · Lindong Wu · Hao Huang · Jinfeng Zhou · Wenchuang Li · Binxin Hu · Wendy Gao · Jiaying Xu · Yiming Liu · Jie Tang · Hongning Wang · Minlie Huang
WikiDO: Evaluating Out-of-Distribution Generalization of Vision-Language Models in Cross-Modal Retrieval
Pavan Kalyan Tankala · Piyush Pasi · Sahil Dharod · Azeem Motiwala · Preethi Jyothi · Aditi Chaudhary · Krishna Srinivasan
GMAI-MMBench: A Comprehensive Multimodal Evaluation Benchmark Towards General Medical AI
pengcheng chen · Jin Ye · Guoan Wang · Yanjun Li · Zhongying Deng · Wei Li · Tianbin Li · Haodong Duan · Ziyang Huang · Yanzhou Su · Benyou Wang · Shaoting Zhang · Bin Fu · Jianfei Cai · Bohan Zhuang · Eric Seibel · Junjun He · Yu Qiao
FedLLM-Bench: Realistic Benchmarks for Federated Learning of Large Language Models
Rui Ye · Rui Ge · Xinyu Zhu · Jingyi Chai · Du Yaxin · Yang Liu · Yanfeng Wang · Siheng Chen
Easy2Hard-Bench: Standardized Difficulty Labels for Profiling LLM Performance and Generalization
Mucong Ding · Chenghao Deng · Jocelyn Choo · Zichu Wu · Aakriti Agrawal · Avi Schwarzschild · Tianyi Zhou · Tom Goldstein · John Langford · Animashree Anandkumar · Furong Huang
NYU CTF Dataset: A Scalable Open-Source Benchmark Dataset for Evaluating LLMs in Offensive Security
Minghao Shao · Sofija Jancheska · Meet Udeshi · Brendan Dolan-Gavitt · haoran xi · Kimberly Milner · Boyuan Chen · Max Yin · Siddharth Garg · Prashanth Krishnamurthy · Farshad Khorrami · Ramesh Karri · Muhammad Shafique
WhodunitBench: Evaluating Large Multimodal Agents via Murder Mystery Games
Junlin Xie · Ruifei Zhang · Zhihong Chen · Xiang Wan · Guanbin Li
MEQA: A Benchmark for Multi-hop Event-centric Question Answering with Explanations
Ruosun Li · Zimu Wang · Son Tran · Lei Xia · Xinya Du
EffiBench: Benchmarking the Efficiency of Automatically Generated Code
Dong HUANG · Yuhao QING · Weiyi Shang · Heming Cui · Jie Zhang
The FineWeb Datasets: Decanting the Web for the Finest Text Data at Scale
Guilherme Penedo · Hynek Kydlíček · Loubna Ben allal · Anton Lozhkov · Margaret Mitchell · Colin Raffel · Leandro Von Werra · Thomas Wolf
VRSBench: A Versatile Vision-Language Benchmark Dataset for Remote Sensing Image Understanding
Xiang Li · Jian Ding · Mohamed Elhoseiny
LINGOLY: A Benchmark of Olympiad-Level Linguistic Reasoning Puzzles in Low Resource and Extinct Languages
Andrew M. Bean · Simeon Hellsten · Harry Mayne · Jabez Magomere · Ethan Chi · Ryan Chi · Scott Hale · Hannah Rose Kirk
VLM4Bio: A Benchmark Dataset to Evaluate Pretrained Vision-Language Models for Trait Discovery from Biological Images
M. Maruf · Arka Daw · Kazi Sajeed Mehrab · Harish Babu Manogaran · Abhilash Neog · Medha Sawhney · Mridul Khurana · James Balhoff · Yasin Bakis · Bahadır Altıntaş · Matthew Thompson · Elizabeth Campolongo · Josef Uyeda · Hilmar Lapp · Henry Bart · Paula Mabee · Yu Su · Wei-Lun (Harry) Chao · Charles Stewart · Tanya Berger-Wolf · Wasila Dahdul · Anuj Karpatne
CVQA: Culturally-diverse Multilingual Visual Question Answering Benchmark
David Romero · Chenyang Lyu · Haryo Wibowo · Santiago Góngora · Aishik Mandal · Sukannya Purkayastha · Jesus-German Ortiz-Barajas · Emilio Cueva · Jinheon Baek · Soyeong Jeong · Injy Hamed · Yong Zheng-Xin · Zheng Wei Lim · Paula Silva · Jocelyn Dunstan · Mélanie Joutiteau · David LE MEUR · Joan Nwatu · Ganzorig Batnasan · Munkh-Erdene Otgonbold · Munkhjargal Gochoo · Guido Ivetta · Luciana Benotti · Laura Alonso Alemany · Hernán Maina · Jiahui Geng · Tiago Timponi Torrent · Frederico Belcavello · Marcelo Viridiano · Jan Christian Blaise Cruz · Dan John Velasco · Oana Ignat · Zara Burzo · Chenxi Whitehouse · Artem Abzaliev · Teresa Clifford · Gráinne Caulfield · Teresa Lynn · Christian Salamea-Palacios · Vladimir Araujo · Yova Kementchedzhieva · Mihail Mihaylov · Israel Azime · Henok Ademtew · Bontu Balcha · Naome A. Etori · David Adelani · Rada Mihalcea · Atnafu Lambebo Tonja · Maria Cabrera · Gisela Vallejo · Holy Lovenia · Ruochen Zhang · Marcos Estecha-Garitagoitia · Mario Rodríguez-Cantelar · Toqeer Ehsan · Rendi Chevi · Muhammad Adilazuarda · Ryandito Diandaru · Samuel Cahyawijaya · Fajri Koto · Tatsuki Kuribayashi · Haiyue Song · Aditya Khandavally · Thanmay Jayakumar · Raj Dabre · Mohamed Imam · Kumaranage Nagasinghe · Alina Dragonetti · Luis Fernando D'Haro · Niyomugisha Olivier · Jay Gala · Pranjal Chitale · Fauzan Farooqui · Tamar Solorio · Alham Aji
FIRE: A Dataset for Feedback Integration and Refinement Evaluation of Multimodal Models
Pengxiang Li · Zhi Gao · Bofei Zhang · Tao Yuan · Yuwei Wu · Mehrtash Harandi · Yunde Jia · Song-Chun Zhu · Qing Li
A Hitchhikers Guide to Fine-Grained Face Forgery Detection Using Common Sense Reasoning

Niki Foteinopoulou · Enjie Ghorbel · Djamila Aouada
GLBench: A Comprehensive Benchmark for Graph with Large Language Models
Yuhan Li · Peisong Wang · Xiao Zhu · Aochuan Chen · Haiyun Jiang · Deng Cai · Wai Kin (Victor) Chan · Jia Li
MedCalc-Bench: Evaluating Large Language Models for Medical Calculations
Nikhil Khandekar · Qiao Jin · Guangzhi Xiong · Soren Dunn · Serina Applebaum · Zain Anwar · Maame Sarfo-Gyamfi · Conrad Safranek · Abid Anwar · Andrew Zhang · Aidan Gilson · Maxwell Singer · Amisha Dave · Anrew Taylor · Aidong Zhang · Qingyu Chen · Zhiyong Lu
COLD: Causal reasOning in cLosed Daily activities
Abhinav Joshi · areeb ahmad · Ashutosh Modi
DevBench: A multimodal developmental benchmark for language learning
Alvin Tan · Chunhua Yu · Bria Long · Wanjing Ma · Tonya Murray · Rebecca Silverman · Jason Yeatman · Michael C Frank
Renovating Names in Open-Vocabulary Segmentation Benchmarks
Haiwen Huang · Songyou Peng · Dan Zhang · Andreas Geiger
Observational Scaling Laws and the Predictability of Language Model Performance
Yangjun Ruan · Chris Maddison · Tatsunori Hashimoto
ConStat: Performance-Based Contamination Detection in Large Language Models
Jasper Dekoninck · Mark Müller · Martin Vechev
Image2Struct: A Benchmark for Evaluating Vision-Language Models in Extracting Structured Information from Images
Josselin Roberts · Tony Lee · Chi Heem Wong · Michihiro Yasunaga · Yifan Mai · Percy Liang
CoLLaM: A Comprehensive Chinese Legal Benchmark for Evaluating Large Language Models
Haitao Li · You Chen · Qingyao Ai · Yueyue WU · Ruizhe Zhang · Yiqun LIU
ConMe: Rethinking Evaluation of Compositional Reasoning for Modern VLMs
Irene Huang · Wei Lin · Muhammad Jehanzeb Mirza · Jacob Hansen · Sivan Doveh · Victor Butoi · Roei Herzig · Assaf Arbelle · Hilde Kuehne · Trevor Darrell · Chuang Gan · Aude Oliva · Rogerio Feris · Leonid Karlinsky
SUGARCREPE++ Dataset: Vision-Language Model Sensitivity to Semantic and Lexical Alterations
Sri Harsha Dumpala · Aman Jaiswal · Chandramouli Shama Sastry · Evangelos Milios · Sageev Oore · Hassan Sajjad
CiteME: Can Language Models Accurately Cite Scientific Claims?
Ori Press · Andreas Hochlehnert · Ameya Prabhu · Vishaal Udandaraao · Ofir Press · Matthias Bethge
InfiBench: Evaluating the Question-Answering Capabilities of Code Large Language Models
Linyi Li · Shijie Geng · Zhenwen Li · Yibo He · Hao Yu · Ziyue Hua · Guanghan Ning · Siwei Wang · Tao Xie · Hongxia Yang
StackEval: Benchmarking LLMs in Coding Assistance
Zulkuf Genc · Nidhish Shah · Dogu Araci
Seshat Global History Databank Text Dataset and Benchmark of Large Language Models' History Knowledge
Jakob Hauser · R. Maria del Rio-Chanona · Dániel Kondor · Majid Benam · Jenny Reddish · Enrico Cioni · Federica Villa · Daniel Hoyer · James Bennett · Pieter Francois · Peter Turchin
SETBENCH: Assessing the Analytical and Semantic Robustness of Language Models
Nicholas Dronen · Bardiya Akhbari · Manish Digambar Gawali
DART-Math: Difficulty-Aware Rejection Tuning for Mathematical Problem-Solving
Yuxuan Tong · Xiwen Zhang · Rui Wang · Ruidong Wu · Junxian He
Beyond Prompts: Dynamic Conversational Benchmarking of Large Language Models
David Castillo-Bolado · Joseph Davidson · Finlay Gray · Marek Rosa
GTA: A Benchmark for General Tool Agents
Jize Wang · Ma Zerun · Yining Li · Songyang Zhang · Caillian Chen · Kai Chen · Xinyi Le
PersonalSum: A User-Subjective Guided Personalized Summarization Dataset for Large Language Models
Lemei Zhang · Peng Liu · Marcus Henriksboe · Even Lauvrak · Jon Atle Gulla · Heri Ramampiaro
BertaQA: How Much Do Language Models Know About Local Culture?
Julen Etxaniz · Gorka Azkune · Aitor Soroa · Oier Lacalle · Mikel Artetxe
ShopBench: A Massive Multi-Task Online Shopping Benchmark for Large Language Models
Yilun Jin · Zheng Li · Chenwei Zhang · Tianyu Cao · Yifan Gao · Pratik Jayarao · Mao Li · Xin Liu · Ritesh Sarkhel · Xianfeng Tang · Haodong Wang · Zhengyang Wang · Wenju Xu · Jingfeng Yang · Qingyu Yin · Xian Li · Priyanka Nigam · Yi Xu · Kai Chen · Qiang Yang · Meng Jiang · Bing Yin
ReMI: A Dataset for Reasoning with Multiple Images
Mehran Kazemi · Nishanth Dikkala · Ankit Anand · Petar Devic · Ishita Dasgupta · Fangyu Liu · Bahare Fatemi · Pranjal Awasthi · Sreenivas Gollapudi · Dee Guo · Ahmed Qureshi
laC-Eval: A code generation benchmark for Infrastructure-as-Code programs
Patrick Tser Jern Kon · Jiachen Liu · Yiming Qiu · Weijun Fan · Ting He · Lei Lin · Haoran Zhang · Owen Park · George Elengikal · Yuxin Kang · Ang Chen · Mosharaf Chowdhury · Myungjin Lee · Xinyu Wang
Beyond Aesthetics: Cultural Competence in Text-to-Image Models
Nithish Kannan Senthilkumar · Arif Ahmad · Marco Andreetto · Vinodkumar Prabhakaran · Utsav Prabhu · Adjai Bousso Dieng · Pushpak Bhattacharyya · Shachi Dave
Benchmarking Trustworthiness of Multimodal Large Language Models: A Comprehensive Study

Yichi Zhang · Yao Huang · Yitong Sun · Chang Liu · Zhe Zhao · Zhengwei Fang · Yifan Wang · Huanran Chen · Xiao Yang · Xingxing Wei · Hang Su · Yinpeng Dong · Jun Zhu
FVEL: Interactive Formal Verification Environment with Large Language Models via Theorem Proving
Xiaohan Lin · Qingxing Cao · Yinya Huang · Haiming Wang · Jianqiao Lu · Zhengying Liu · Linqi Song · Xiaodan Liang
ClevrSkills: Compositional Language And Visual Understanding in Robotics
Sanjay Haresh · Daniel Dijkman · Apratim Bhattacharyya · Roland Memisevic
Are Large Language Models Good Statisticians?
Yizhang Zhu · Shiyin Du · Boyan Li · Yuyu Luo · Nan Tang
ERBench: An Entity-Relationship based Automatically Verifiable Hallucination Benchmark for Large Language Models
Jio Oh · Soyeon Kim · Junseok Seo · Jindong Wang · Ruochen Xu · Xing Xie · Steven Whang
JailbreakBench: An Open Robustness Benchmark for Jailbreaking Large Language Models
Patrick Chao · Edoardo Debenedetti · Alexander Robey · Maksym Andriushchenko · Francesco Croce · Vikash Sehwag · Edgar Dobriban · Nicolas Flammarion · George J. Pappas · Florian Tramer · Hamed Hassani · Eric Wong
Evaluate calibration of language models with folktexts
André F. Cruz · Celestine Mender-Dünner · Moritz Hardt
Me, Myself, and AI: The Situational Awareness Dataset (SAD) for LLMs
Rudolf Laine · Bilal Chughtai · Jan Betley · Kaivalya Hariharan · Mikita Balesni · Jérémy Scheurer · Marius Hobbhahn · Alexander Meinke · Owain Evans
NewTerm: Benchmarking Real-Time New Terms for Large Language Models with Annual Updates
Hexuan Deng · Wenxiang Jiao · Xuebo Liu · Min Zhang · Zhaopeng Tu
MedJourney: Benchmark and Evaluation of Large Language Models over Patient Clinical Journey
Xian Wu · Yutian Zhao · Yunyan Zhang · Jiageng Wu · Zhihong Zhu · Yingying Zhang · Yi Ouyang · Ziheng Zhang · Huimin WANG · zhenxi Lin · Jie Yang · Shuang Zhao · Yefeng Zheng
Revisiting Few-Shot Object Detection with Vision-Language Models
Anish Madan · Neehar Peri · Shu Kong · Deva Ramanan
SD-Eval: A Benchmark Dataset for Spoken Dialogue Understanding Beyond Words
Junyi AO · Yuancheng Wang · Xiaohai Tian · Dekun Chen · Jun Zhang · Lu Lu · Yuxuan Wang · Haizhou Li · Zhizheng Wu
MathPile: A Billion-Token-Scale Pretraining Corpus for Math
Zengzhi Wang · Xuefeng Li · Rui Xia · Pengfei Liu
A SARS-CoV-2 Interaction Dataset and VHH Sequence Corpus for Antibody Language Models
Hirofumi Tsuruta · Hiroyuki Yamazaki · Ryota Maeda · Ryotaro Tamura · Akihiro Imura
DataComp-LM: In search of the next generation of training sets for language models
Amro Abbas · Alon Albalak · Kushal Arora · Hritik Bansal · Yonatan Bitton · Yair Carmon · Khyathi Chandu · Mayee Chen · Giannis Daras · Achal Dave · Alex Dimakis · Alaaeldin El-Nouby · Fartash Faghri · Alex Fang · Samir Yitzhak Gadre · Josh Gardner · Saurabh Garg · Dhruva Ghosh · Aaron Gokaslan · Dirk Groeneveld · Etash Guha · Suchin Gururangan · Reinhard Heckel · Cheng-Yu Hsieh · Gabriel Ilharco · Maor Ivgi · Jenia Jitsev · Matt Jordan · Sham Kakade · Sedrick Scott Keh · Maciej Kilian · Pang Wei Koh · Thomas Kollar · Jeffrey Li · Kyle Lo · Kalyani Marathe · Jean Mercat · Niklas Muennighoff · Marianna Nezhurina · Thao Nguyen · Sewoong Oh · Hadi Pouransari · Sarah Pratt · Sunny Sanyal · Ludwig Schmidt · Vaishaal Shankar · Rulin Shao · Georgios Smyrnis · Luca Soldaini · Shuran Song · Alexander Toshev · Igor Vasiljevic · Stephanie Wang · Mitchell Wortsman · Rui Xin · Luke Zettlemoyer · Hanlin Zhang · Jieyu Zhang
UniBench: Visual Reasoning Requires Rethinking Vision-Language Beyond Scaling
Haider Al-Tahan · Quentin Garrido · Randall Balestriero · Diane Bouchacourt · Caner Hazirbas · Mark Ibrahim
JourneyBench: A Challenging One-Stop Vision-Language Understanding Benchmark of Generated Images
Zhecan Wang · Junzhang Liu · Chia-Wei Tang · Hani Alomari · Anushka Sivakumar · Rui Sun · Wenhao Li · Md. Atabuzzaman · Hammad Ayyubi · Haoxuan You · Alvi Md Ishmam · Kai-Wei Chang · Shih-Fu Chang · Christopher Thomas
TEG-DB: A Comprehensive Dataset and Benchmark of Textual-Edge Graphs
Zhuofeng Li · Zixing Gou · Xiangnan Zhang · Zhongyuan Liu · Sirui Li · Yuntong Hu · Chen LING · Zheng Zhang · Liang Zhao
Instruction Tuning Large Language Models to Understand Electronic Health Records
Zhenbang Wu · Anant Dadu · Michael Nalls · Faraz Faghri · Jimeng Sun
CARES: A Comprehensive Benchmark of Trustworthiness in Medical Vision Language Models
Peng Xia · Ze Chen · Juanxi Tian · Yangrui Gong · Ruibo Hou · Yue Xu · Zhenbang Wu · Zhiyuan Fan · Yiyang Zhou · Kangyu Zhu · Wenhao Zheng · Zhaoyang Wang · Xiao Wang · Xuchao Zhang · Chetan Bansal · Marc Niethammer · Junzhou Huang · Hongtu Zhu · Yun Li · Jimeng Sun · Zongyuan Ge · Gang Li · James Zou · Huaxiu Yao
What to Say and When to Say it: Live Fitness Coaching as a Testbed for Situated Interaction
Sunny Panchal · Apratim Bhattacharyya · Guillaume Berger · Antoine Mercier · Cornelius Böhm · Florian Dietrichkeit · Reza Pourreza · Xuanlin Li · Pulkit Madan · Mingu Lee · Mark Todorovich · Ingo Bax · Roland Memisevic
CompBench: A Comparative Reasoning Benchmark for Multimodal LLMs
Jihyung Kil · Zheda Mai · Justin Lee · Zihe Wang · Kerrie Cheng · Lemeng Wang · Ye Liu · Arpita Chowdhury · Wei-Lun (Harry) Chao
STaRK: Benchmarking LLM Retrieval on Textual and Relational Knowledge Bases

Shirley Wu · Shiyu Zhao · Michihiro Yasunaga · Kexin Huang · Kaidi Cao · Qian Huang · Vassilis Ioannidis · Karthik Subbian · James Zou · Jure Leskovec
Web2Code: A Large-scale Webpage-to-Code Dataset and Evaluation Framework for Multimodal LLMs
Sukmin Yun · haokun lin · Rusiru Thushara · Mohammad Bhat · Yongxin Wang · zutao jiang · Mingkai Deng · Jinhong Wang · Tianhua Tao · Junbo Li · Haonan Li · Preslav Nakov · Timothy Baldwin · Zhengzhong Liu · Eric Xing · Xiaodan Liang · Zhiqiang Shen
Paloma: A Benchmark for Evaluating Language Model Fit
Ian Magnusson · Akshita Bhagia · Valentin Hofmann · Luca Soldaini · Ananya Harsh Jha · Oyvind Tafjord · Dustin Schwenk · Evan Walsh · Yanai Elazar · Kyle Lo · Dirk Groeneveld · Iz Beltagy · Hannaneh Hajishirzi · Noah Smith · Kyle Richardson · Jesse Dodge
Who Evaluates the Evaluations? Objectively Scoring Text-to-Image Prompt Coherence Metrics with T2IScoreScore (TS2)
Michael Saxon · Fatima Jahara · Mahsa Khoshnoodi · Yujie Lu · Aditya Sharma · William Yang Wang
LLMCBench: Benchmarking Large Language Model Compression for Efficient Deployment
Jinyang Guo · Ge Yang · Changyi He · Jianyu Wu · Yifu Ding · Aishan Liu · Haotong Qin · Pengliang Ji · Xianglong Liu
MMDU: A Multi-Turn Multi-Image Dialog Understanding Benchmark and Instruction-Tuning Dataset for LVLs
Ziyu Liu · Tao Chu · Yuhang Zang · Xilin Wei · Xiaoyi Dong · Pan Zhang · Zijian Liang · Yuanjun Xiong · Dahua Lin · Yu Qiao · Jiaqi Wang
EvoCodeBench: An Evolving Code Generation Benchmark with Domain-Specific Evaluations
Jia Li · Ge Li · Xuanming Zhang · YunFei Zhao · Yihong Dong · Zhi Jin · Binhua Li · Fei Huang · Yongbin Li
WildTeaming at Scale: From In-the-Wild Jailbreaks to (Adversarially) Safer Language Models
Liwei Jiang · Kavel Rao · Seungju Han · Allyson Ettinger · Faeze Brahman · Sachin Kumar · Niloofar Mireshghallah · Ximing Lu · Maarten Sap · Yejin Choi · Nouha Dziri
CoIN: A Benchmark of Continual Instruction Tuning for Multimodal Large Language Models
Cheng Chen · Junchen Zhu · Xu Luo · Hengtao Shen · Lianli Gao · Jingkuan Song
Newswire: A Large-Scale Structured Database of a Century of Historical News
Emily Silcock · Abhishek Arora · Luca D'Amico-Wong · Melissa Dell
\$\texttt{ConflictBank}\$: A Benchmark for Evaluating the Influence of Knowledge Conflicts in LLMs
Zhaochen Su · Jun Zhang · Xiaoye Qu · Tong Zhu · Yanshu Li · Jiashuo Sun · Juntao Li · Min Zhang · Yu Cheng
Weak-eval-Strong: Evaluating and Eliciting Lateral Thinking of LLMs with Situation Puzzles
Qi Chen · Bowen Zhang · Gang Wang · Qi Wu
Task Me Anything
Jieyu Zhang · Weikai Huang · Zixian Ma · Oscar Michel · Dong He · Tanmay Gupta · Wei-Chiu Ma · Ali Farhadi · Aniruddha Kembhavi · Ranjay Krishna
Benchmarking LLMs via Uncertainty Quantification
Fanghua Ye · Mingming Yang · Jianhui Pang · Longyue Wang · Derek Wong · Emine Yilmaz · Shuming Shi · Zhaopeng Tu
II-Bench: An Image Implication Understanding Benchmark for Multimodal Large Language Models
Ziqiang Liu · Feiteng Fang · Xi Feng · Xeron Du · Chenhao Zhang · Noah Wang · yuelin bai · Qixuan Zhao · Liyang Fan · CHENGQUANG GAN · Hongquan Lin · Jiaming Li · Yuansheng Ni · Haihong Wu · Yaswanth Narsupalli · Zhigang Zheng · Chengming Li · Xiping Hu · Ruifeng Xu · Xiaojun Chen · Min Yang · Jiaheng Liu · Ruibo Liu · Wenhao Huang · Ge Zhang · Shiwen Ni
MMBench-Video: A Long-Form Multi-Shot Benchmark for Holistic Video Understanding
Xinyu Fang · Kangrui Mao · Haodong Duan · Xiangyu Zhao · Yining Li · Dahua Lin · Kai Chen
RepLiQA: A Question-Answering Dataset for Benchmarking LLMs on Unseen Reference Content
Joao Monteiro · Pierre-André Noël · Étienne Marcotte · Sai Rajeswar Mudumba · Valentina Zantedeschi · David Vazquez · Nicolas Chapados · Chris Pal · Perouz Taslakian
Are We on the Right Way for Evaluating Large Vision-Language Models?
Lin Chen · Jinsong Li · Xiaoyi Dong · Pan Zhang · Yuhang Zang · Zehui Chen · Haodong Duan · Jiaqi Wang · Yu Qiao · Dahua Lin · Feng Zhao
WenMind: A Comprehensive Benchmark for Evaluating Large Language Models in Chinese Classical Literature and Language Arts
Jiahuan Cao · Yang Liu · Yongxin Shi · Kai Ding · Lianwen Jin
CriticEval: Evaluating Large-scale Language Model as Critic
Tian Lan · Wenwei Zhang · Chen Xu · Heyan Huang · Dahua Lin · Kai Chen · Xian-Ling Mao
MMM-RS: A Multi-modal, Multi-GSD, Multi-scene Remote Sensing Dataset and Benchmark for Text-to-Image Generation
jialin luo · Yuanzhi Wang · Ziqi Gu · Yide Qiu · Shuaizhen Yao · Fuyun Wang · Chunyan Xu · Wenhua Zhang · Dan Wang · Zhen Cui
RWKU: Benchmarking Real-World Knowledge Unlearning for Large Language Models
Zhuoran Jin · Pengfei Cao · Chenhao Wang · Zhitao He · Hongbang Yuan · Jiachun Li · Yubo Chen · Kang Liu · Jun Zhao
E.T. Bench: Towards Open-Ended Event-Level Video-Language Understanding
Ye Liu · Zongyang Ma · Zhongang Qi · Yang Wu · Ying Shan · Chang Chen
BABILong: Testing the Limits of LLMs with Long Context Reasoning-in-a-Haystack
Yury Kuratov · Aydar Bulatov · Petr Anokhin · Ivan Rodkin · Dmitry Sorokin · Artyom Sorokin · Mikhail Burtsev

CableInspect-AD: An Expert-Annotated Anomaly Detection Dataset

Akshatha Arodi · Margaux Luck · Jean-Luc Bedwani · Aldo Zaimi · Ge Li · Nicolas Pouliot · Julien Beaudry · Gaetan Marceau Caron

Data Generation**Carrot and Stick: Eliciting Comparison Data and Beyond**

Yiling Chen · Shi Feng · Fang-Yi Yu

Partially Observable Cost-Aware Active-Learning with Large Language Models

Nicolás Astorga · Tennison Liu · Nabeel Seedat · Mihaela van der Schaar

GAVEL: Generating Games via Evolution and Language Models

Graham Todd · Alexander G Padula · Matthew Stephenson · Eric Piette · Dennis Soemers · Julian Togelius

NoiseGPT: Label Noise Detection and Rectification through Probability Curvature

Haoyu Wang · Zhuo Huang · Zhiwei Lin · Tongliang Liu

CLUES: Collaborative Private-domain High-quality Data Selection for LLMs via Training Dynamics

Wanru Zhao · Hongxiang Fan · Shell Xu Hu · Wangchunshu Zhou · Nicholas Lane

ACES: Generating a Diversity of Challenging Programming Puzzles with Autotelic Generative Models

Julien Pourcel · Cédric Colas · Gaia Molinaro · Pierre-Yves Oudeyer · Laetitia Teodorescu

EPIC: Effective Prompting for Imbalanced-Class Data Synthesis in Tabular Data Classification via Large Language Models

Jinhee Kim · Taesung Kim · Jaegul Choo

The ALCHEmist: Automated Labeling 500x CHEaper than LLM Data Annotators

Tzu-Heng Huang · Catherine Cao · Vaishnavi Bhargava · Frederic Sala

End-To-End Causal Effect Estimation from Unstructured Natural Language Data

Nikita Dhawan · Leonardo Cotta · Karen Ullrich · Rahul Krishnan · Chris Maddison

cPAPERS: A Dataset of Situated and Multimodal Interactive Conversations in Scientific Papers

Anirudh Sundar · Jin Xu · William Gay · Christopher Richardson · Larry Heck

Neuro-Symbolic Data Generation for Math Reasoning

Zenan Li · Zhi Zhou · Yuan Yao · Xian Zhang · Yu-Feng Li · Chun Cao · Fan Yang · Xiaoxing Ma

EvolveDirector: Approaching Advanced Text-to-Image Generation with Large Vision-Language Models

Rui Zhao · Hangjie Yuan · Yujie Wei · Shiwei Zhang · Yuchao Gu · Lingmin Ran · Xiang Wang · Jay Zhangjie Wu · David Junhao Zhang · Yingya Zhang · Mike Zheng Shou

Lean Workbook: A large-scale Lean problem set formalized from natural language math problems

Huaiyuan Ying · Zijian Wu · Yihan Geng · Jlayu Wang · Dahua Lin · Kai Chen

LLM-AutoDA: Large Language Model-Driven Automatic Data Augmentation for Long-tailed Problems

Pengkun Wang · Zhe Zhao · HaiBin Wen · Fanfu Wang · Binwu Wang · Qingfu Zhang · Yang Wang

Efficient LLM**Efficient Contextual LLM Cascades through Budget-Constrained Policy Learning**

Xuechen Zhang · Zijian Huang · Ege Onur Taga · Carlee Joe-Wong · Samet Oymak · Jiasi Chen

AMAGO-2: Breaking the Multi-Task Barrier in Meta-Reinforcement Learning with Transformers

Jake Grigsby · Justin Sasek · Samyak Parajuli · Ikechukwu D. Adebisi · Amy Zhang · Yuke Zhu

Cherry on Top: Parameter Heterogeneity and Quantization in Large Language Models

Wanyun Cui · Qianle Wang

An Efficient Recipe for Long Context Extension via Middle-Focused Positional Encoding

Tong Wu · Yanpeng Zhao · Zilong Zheng

Thinking Forward: Memory-Efficient Federated Finetuning of Language Models

Kunjal Panchal · Nisarg Parikh · Sunav Choudhary · Lijun Zhang · Yuriy Brun · Hui Guan

Searching for Efficient Linear Layers over a Continuous Space of Structured Matrices

Andres Potapczynski · Shikai Qiu · Marc Finzi · Christopher Ferri · Charlie Chen · Micah Goldblum · C. Bayan Bruss · Christopher De Sa · Andrew Wilson

Dataset Decomposition: Faster LLM Training with Variable Sequence Length Curriculum

Hadi Pouransari · Chun-Liang Li · Jen-Hao Chang · Pavan Kumar Anasosalu Vasu · Cem Koc · Vaishaal Shankar · Orel Tuzel

Sequoia: Scalable and Robust Speculative Decoding

Zhuoming Chen · Avner May · Ruslan Svirschevski · Yu-Hsun Huang · Max Ryabinin · Zhihao Jia · Beidi Chen

Reversing the Forget-Retain Objectives: An Efficient LLM Unlearning Framework from Logit Difference

Jiabao Ji · Yujian Liu · Yang Zhang · Gaowen Liu · Ramana Kompella · Sijia Liu · Shiyu Chang

Training Dynamics of Transformers to Recognize Word Co-occurrence via Gradient Flow Analysis

Hongru Yang · Bhavya Kailkhura · Zhangyang “Atlas” Wang · Yingbin Liang

3-in-1: 2D Rotary Adaptation for Efficient Finetuning, Efficient Batching and Composability

Baohao Liao · Christof Monz

Be like a Goldfish, Don't Memorize! Mitigating Memorization in Generative LLMs

Abhimanyu Hans · John Kirchenbauer · Yuxin Wen · Neel Jain · Hamid Kazemi · Prajwal Singhania · Siddharth Singh · Gowthami Somepalli · Jonas Geiping · Abhinav Bhatele · Tom Goldstein

B'MOJO: Hybrid State Space Realizations of Foundation Models with Eidetic and Fading Memory
Luca Zancato · Arjun Seshadri · Yonatan Dukler · Aditya Sharad Golatkar · Yantao Shen · Benjamin Bowman · Matthew Trager · Alessandro Achille · Stefano Soatto
Fundamental Limits of Prompt Compression: A Rate-Distortion Framework for Black-Box Language Models
Alliot Nagle · Adway Girish · Marco Bondaschi · Michael Gastpar · Ashok Vardhan Makkuva · Hyeji Kim
Building on Efficient Foundations: Effective Training of LLMs with Structured Feedforward Layers
Xiuying Wei · Skander Moalla · Razvan Pascanu · Caglar Gulcehre
Scaling Laws with Vocabulary: Larger Models Deserve Larger Vocabularies
Chaofan Tao · Qian Liu · Longxu Dou · Niklas Muennighoff · Zhongwei Wan · Ping Luo · Min Lin · Ngai Wong
DropBP: Accelerating Fine-Tuning of Large Language Models by Dropping Backward Propagation
Sunghyeon Woo · Baeseong Park · Byeongwook Kim · Minjung Jo · Se Jung Kwon · Dongsuk Jeon · Dongsoo Lee
KV Cache is 1 Bit Per Channel: Efficient Large Language Model Inference with Coupled Quantization
Tianyi Zhang · Jonah Yi · Zhaozhuo Xu · Anshumali Shrivastava
Twin-Merging: Dynamic Integration of Modular Expertise in Model Merging
Zhenyi Lu · Chenghao Fan · Wei Wei · Xiaoye Qu · Dangyang Chen · Yu Cheng
Accelerating Greedy Coordinate Gradient and General Prompt Optimization via Probe Sampling
Yiran Zhao · Wenye Zheng · Tianle Cai · Do Xuan Long · Kenji Kawaguchi · Anirudh Goyal · Michael Qizhe Shieh
Accuracy is Not All You Need
Abhinav Dutta · Sanjeev Krishnan · Nipun Kwatra · Ramachandran Ramjee
SDP4Bit: Toward 4-bit Communication Quantization in Sharded Data Parallelism for LLM Training
Jinda Jia · Cong Xie · Hanlin Lu · Daoce Wang · Hao Feng · Chengming Zhang · Baixi Sun · Haibin Lin · Zhi Zhang · Xin Liu · Dingwen Tao
DeeR-VLA: Dynamic Inference of Multimodal Large Language Models for Efficient Robot Execution
Yang Yue · Yulin Wang · Bingyi Kang · Yizeng Han · Shenzi Wang · Shiji Song · Jiashi Feng · Gao Huang
Transformers to SSMS: Distilling Quadratic Knowledge to Subquadratic Models
Aviv Bick · Kevin Li · Eric Xing · J. Zico Kolter · Albert Gu
OccamLLM: Fast and Exact Language Model Arithmetic in a Single Step
Owen Dugan · Donato Jiménez-Beneto · Charlotte Loh · Zhuo Chen · Rumen Dangovski · Marin Soljacic
Toward Efficient Inference for Mixture of Experts
Haiyang Huang · Newsha Ardalani · Anna Sun · Liu Ke · Shruti Bhosale · Hsien-Hsin Lee · Carole-Jean Wu · Benjamin Lee
Enhancing LLM's Cognition via Structurization
Kai Liu · Zhihang Fu · Chao Chen · Wei Zhang · Rongxin Jiang · Fan Zhou · Yaowu Chen · Yue Wu · Jieping Ye
Mixture of Scales: Memory-Efficient Token-Adaptive Binarization for Large Language Models
Dongwon Jo · Taesu Kim · Yulhwa Kim · jae-joon kim
Predictor-Corrector Enhanced Transformers with Exponential Moving Average Coefficient Learning
Bei Li · Tong Zheng · Rui Wang · Jiahao Liu · 清妍 郭 · Junliang Guo · Xu Tan · Tong Xiao · JingBo Zhu · Jingang Wang · Xunliang Cai
Normalization Layer Per-Example Gradients are Sufficient to Predict Gradient Noise Scale in Transformers
Gavia Gray · aman tiwari · Shane Bergsma · Joel Hestness
Unlocking Tokens as Data Points for Generalization Bounds on Larger Language Models
Sanae Lotfi · Yilun Kuang · Marc Finzi · Brandon Amos · Micah Goldblum · Andrew Wilson
Tiny Time Mixers (TTMs): Fast Pre-trained Models for Enhanced Zero/Few-Shot Forecasting of Multivariate Time Series
Vijay Ekambaram · Arindam Jati · Pankaj Dayama · Sumanta Mukherjee · Nam Nguyen · Wesley M Gifford · Chandra Reddy · Jayant Kalagnanam
Enhancing Large Language Models through Adaptive Tokenizers
Mengyu Zheng · Hanting Chen · Tianyu Guo · Chong Zhu · Binfan Zheng · Chang Xu · Yunhe Wang
DISP-LLM: Dimension-Independent Structural Pruning for Large Language Models
Shangqian Gao · Chi-Heng Lin · Ting Hua · Zheng Tang · Yilin Shen · Hongxia Jin · Yen-Chang Hsu
SpaceByte: Towards Deleting Tokenization from Large Language Modeling
Kevin Slagle
AmoebaLLM: Constructing Any-Shape Large Language Models for Efficient and Instant Deployment
Yonggan Fu · Zhongzhi Yu · Junwei Li · Jiayi Qian · Yongan Zhang · Xiangchi Yuan · Dachuan Shi · Roman Yakunin · Yingyan (Celine) Lin
BAM! Just Like That: Simple and Efficient Parameter Upcycling for Mixture of Experts
Qizhen (Irene) Zhang · Nikolas Gritsch · Dwaraknath Gnaneshwar Talupuru · Simon Guo · David Cairuz · Bharat Venkitesh · Jakob Foerster · Phil Blunsom · Sebastian Ruder · Ahmet Üstün · Acyr Locatelli
Orchid: Flexible and Data-Dependent Convolution for Sequence Modeling
Mahdi Karami · Ali Ghodsi
PV-Tuning: Beyond Straight-Through Estimation for Extreme LLM Compression
Vladimir Malinovskii · Denis Mazur · Ivan Ilin · Denis Kuznedelev · Konstantin Burlachenko · Kai Yi · Dan Alistarh · Peter Richtarik
Improving Deep Learning Optimization through Constrained Parameter Regularization

Jörg Franke · Michael Hefenbrock · Gregor Koehler · Frank Hutter
Cross-model Control: Improving Multiple Large Language Models in One-time Training
Jiayi Wu · Hao Sun · Hengyi Cai · Lixin Su · Shuaiqiang Wang · Dawei Yin · Xiang Li · Ming Gao
Multi-Head Mixture-of-Experts
Xun Wu · Shaohan Huang · Wenhui Wang · Shuming Ma · Li Dong · Furu Wei
Simple and Effective Masked Diffusion Language Models
Subham Sahoo · Marianne Arriola · Aaron Gokaslan · Edgar Marroquin · Alexander Rush · Yair Schiff · Justin Chiu · Volodymyr Kuleshov
Weight decay induces low-rank attention layers
Seijin Kobayashi · Yassir Akram · Johannes von Oswald
Thought of Search: Planning with Language Models Through The Lens of Efficiency
Michael Katz · Harsha Kokel · Kavitha Srinivas · Shirin Sohrabi Araghi
Mixture of In-Context Experts Enhance LLMs' Long Context Awareness
Hongzhan Lin · Ang Lv · yuhan chen · chen zhu · Yang Song · Hengshu Zhu · Rui Yan
ProTransformer: Robustify Transformers via Plug-and-Play Paradigm
Zhichao Hou · Weizhi Gao · Yuchen Shen · Feiyi Wang · Xiaorui Liu
Learn To be Efficient: Build Structured Sparsity in Large Language Models
Haizhong Zheng · Xiaoyan Bai · Xueshen Liu · Zhuoqing Morley Mao · Beidi Chen · Fan Lai · Atul Prakash
Compressing Large Language Models using Low Rank and Low Precision Decomposition
Rajarshi Saha · Naomi Sagan · Varun Srivastava · Andrea Goldsmith · Mert Pilanci
SLTrain: a sparse plus low rank approach for parameter and memory efficient pretraining
Andi Han · Jiaxiang Li · Wei Huang · Mingyi Hong · Akiko Takeda · Pratik Kumar Jawanpuria · Bamdev Mishra
Learning and Transferring Sparse Contextual Bigrams with Linear Transformers
Yunwei Ren · Zixuan Wang · Jason Lee
SwitchHead: Accelerating Transformers with Mixture-of-Experts Attention
Róbert Csordás · Piotr Piękos · Kazuki Irie · Jürgen Schmidhuber
VideoLLM-MoD: Efficient Video-Language Streaming with Mixture-of-Depths Vision Computation
Shiwei Wu · Joya Chen · Kevin Qinghong Lin · Qimeng Wang · Yan Gao · Qianli Xu · Tong Xu · Yao Hu · Enhong Chen · Mike Zheng Shou
Zero-Shot Tokenizer Transfer
Benjamin Minixhofer · Edoardo Maria Ponti · Ivan Vulić
Rethinking Memory and Communication Costs for Efficient Data Parallel Training of Large Language Models
Hanxiao Zhang · Lin JU · Chan Wu · Jinjing Huang · Youshao Xiao · Zhenglei Zhou · Zhiming fan · Zhaoxin Huan · Siyuan Li · Fanzhuang Meng · Lei Liang · Xiaolu Zhang · Jun Zhou
Mesa-Extrapolation: A Weave Position Encoding Method for Enhanced Extrapolation in LLMs
Xin Ma · Yang Liu · Jingjing Liu · Xiaoxu Ma
ReLIZO: Sample Reusable Linear Interpolation-based Zeroth-order Optimization
Xiaoxing Wang · Xiaohan Qin · Xiaokang Yang · Junchi Yan
Implicit Regularization of Sharpness-Aware Minimization for Scale-Invariant Problems
Bingcong Li · Liang Zhang · Niao He
Pipeline Parallelism with Controllable Memory
Penghui Qi · Xinyi Wan · Nyamdavaa Amar · Min Lin
Personalized Adapter for Large Meteorology Model on Devices: Towards Weather Foundation Models
Shengchao Chen · Guodong Long · Jing Jiang · Chengqi Zhang
Q-VLM: Post-training Quantization for Large Vision-Language Models
Changyuan Wang · Ziwei Wang · Xiuwei Xu · Yansong Tang · Jie Zhou · Jiwen Lu
StreamingDialogue: Prolonged Dialogue Learning via Long Context Compression with Minimal Losses
JIANAN LI · Quan Tu · Cunli Mao · Zhengtao Yu · Ji-Rong Wen · Rui Yan
InfLLM: Training-Free Long-Context Extrapolation for LLMs with an Efficient Context Memory
Chaojun Xiao · Pengle Zhang · Xu Han · Guangxuan Xiao · Yankai Lin · Zhengyan Zhang · Zhiyuan Liu · Maosong Sun
VeLoRA: Memory Efficient Training using Rank-1 Sub-Token Projections
Roy Miles · Pradyumna Reddy · Ismail Elezi · Jiankang Deng
SGLang: Efficient Execution of Structured Language Model Programs
Lianmin Zheng · Liangsheng Yin · Zhiqiang Xie · Chuyue (Livia) Sun · Jeff Huang · Cody Hao Yu · Shiyi Cao · Christos Kozyrakis · Ion Stoica · Joseph Gonzalez · Clark Barrett · Ying Sheng
D-LLM: A Token Adaptive Computing Resource Allocation Strategy for Large Language Models
Yikun Jiang · Huanyu Wang · Lei Xie · Hanbin Zhao · zhang chao · Hui Qian · John C.S. Lui
ArkVale: Efficient Generative LLM Inference with Recallable Key-Value Eviction
Renze Chen · Zhuofeng Wang · Beiquan Cao · Tong Wu · Size Zheng · Xiuhong Li · Xuechao Wei · Shengen Yan · Meng Li · Yun Liang
FM-Delta: Lossless Compression for Storing Massive Fine-tuned Foundation Models

Wanyi Ning · Jingyu Wang · Qi Qi · Mengde Zhu · Haifeng Sun · Daixuan Cheng · Jianxin Liao · Ce Zhang
AutoMix: Automatically Mixing Language Models
Pranjal Aggarwal · Aman Madaan · Ankit Anand · Srividya Pranavi Potharaju · Swaroop Mishra · Pei Zhou · Aditya Gupta · Dheeraj Rajagopal · Karthik Kappaganthu · Yiming Yang · Shyam Upadhyay · Manaal Faruqui · Mausam
Efficient Multi-task LLM Quantization and Serving for Multiple LoRA Adapters
Yifei Xia · Fangcheng Fu · Wentao Zhang · Jiawei Jiang · Bin CUI
Heavy-Tailed Class Imbalance and Why Adam Outperforms Gradient Descent on Language Models
Frederik Kunstner · Robin Yadav · Alan Milligan · Mark Schmidt · Alberto Bietti
LoQT: Low-Rank Adapters for Quantized Pretraining
Sebastian Loeschcke · Mads Toftrup · Michael Kastoryano · Serge Belongie · Vésteinn Snæbjarnarson
Accelerating Blockwise Parallel Language Models with Draft Refinement
Taehyeon Kim · Ananda Theertha Suresh · Kishore Papineni · Michael D Riley · Sanjiv Kumar · Adrian Benton
Elliptical Attention
Stefan Nielsen · Laziz Abdullaev · Rachel S.Y. Teo · Tan Nguyen
DARG: Dynamic Evaluation of Large Language Models via Adaptive Reasoning Graph
Zhehao Zhang · Jiaao Chen · Diyi Yang
Adaptive Layer Sparsity for Large Language Models via Activation Correlation Assessment
Wei Li · Lujun Li · Mark Lee · Shengjie Sun
Model-GLUE\$: Democratized LLM Scaling for A Large Model Zoo in the Wild
Xinyu Zhao · Guoheng Sun · Ruisi Cai · Yukun Zhou · Pingzhi Li · Peihao Wang · Bowen Tan · Yexiao He · Li Chen · Yi Liang · Beidi Chen · Binhang Yuan · Hongyi Wang · Ang Li · Zhangyang “Atlas” Wang · Tianlong Chen
Discovering Sparsity Allocation for Layer-wise Pruning of Large Language Models
Lujun Li · Peijie Dong · Zhenheng Tang · Xiang Liu · Qiang Wang · Wenhan Luo · Wei Xue · Qifeng Liu · Xiaowen Chu · Yike Guo
GREATS: Online Selection of High-Quality Data for LLM Training in Every Iteration
Jiachen (Tianhao) Wang · Tong Wu · Dawn Song · Prateek Mittal · Ruoxi Jia
Repurposing Language Models into Embedding Models: Finding the Compute-Optimal Recipe
Albert Q. Jiang · Alicja Ziarko · Bartosz Piotrowski · Wenda Li · Mateja Jamnik · Piotr Miłoś
SpeedLoader: An I/O efficient scheme for heterogeneous and distributed LLM operation
Yiqi Zhang · Yang You
Exploring Context Window of Large Language Models via Decomposed Positional Vectors
Zican Dong · Junyi Li · Xin Men · Xin Zhao · Bingning Wang · Zhen Tian · weipeng chen · Ji-Rong Wen
SnapKV: LLM Knows What You are Looking for Before Generation
Yuhong Li · Yingbing Huang · Bowen Yang · Bharat Venkitesh · Acyr Locatelli · Hanchen Ye · Tianle Cai · Patrick Lewis · Deming Chen
Time-FFM: Towards LM-Empowered Federated Foundation Model for Time Series Forecasting
Qingxiang Liu · Xu Liu · Chenghao Liu · Qingsong Wen · Yuxuan Liang
SpecExec: Massively Parallel Speculative Decoding For Interactive LLM Inference on Consumer Devices
Ruslan Svirschevski · Avner May · Zhuoming Chen · Beidi Chen · Zhihao Jia · Max Ryabinin
ShiftAddLLM: Accelerating Pretrained LLMs via Post-Training Multiplication-Less Reparameterization
Haoran You · Yipin Guo · Yichao Fu · Wei Zhou · Huihong Shi · Xiaofan Zhang · Souvik Kundu · Amir Yazdanbakhsh · Yingyan (Celine) Lin
Efficient LLM Scheduling by Learning to Rank
Yichao Fu · Siqi Zhu · Runlong Su · Aurick Qiao · Ion Stoica · Hao Zhang
Kraken: Inherently Parallel Transformers For Efficient Multi-Device Inference
Rohan Baskar Prabhakar · Hengrui Zhang · David Wentzlaff
EffiLearner: Enhancing Efficiency of Generated Code via Self-Optimization
Dong HUANG · Jianbo Dai · Han Weng · Puzhen Wu · Yuhao QING · Heming Cui · Zhijiang Guo · Jie Zhang
4-bit Shampoo for Memory-Efficient Network Training
Sike Wang · Pan Zhou · Jia Li · Hua Huang
Towards Next-Level Post-Training Quantization of Hyper-Scale Transformers
Junhan Kim · Chungman Lee · Eulrang Cho · Kyungphil Park · Ho-young Kim · Joonyoung Kim · Yongkweon Jeon
A Bayesian Approach to Data Point Selection
XINNUO XU · Minyoung Kim · Royson Lee · Brais Martinez · Timothy Hospedales
Compact Language Models via Pruning and Knowledge Distillation
Saurav Muralidharan · Sharath Turuvekere Sreenivas · Raviraj Joshi · Marcin Chochowski · Mostofa Patwary · Mohammad Shoeybi · Bryan Catanzaro · Jan Kautz · Pavlo Molchanov
Resolving Discrepancies in Compute-Optimal Scaling of Language Models
Tomer Porian · Mitchell Wortsman · Jenia Jitsev · Ludwig Schmidt · Yair Carmon
The Fine-Grained Complexity of Gradient Computation for Training Large Language Models
Josh Alman · Zhao Song
SparseLLM: Towards Global Pruning of Pre-trained Language Models

Guangji Bai · Yijiang Li · Chen LING · Kibaek Kim · Liang Zhao
Found in the Middle: How Language Models Use Long Contexts Better via Plug-and-Play Positional Encoding
Zhenyu Zhang · Runjin Chen · Shiwei Liu · Zhewei Yao · Olatunji Ruwase · Beidi Chen · Xiaoxia Wu · Zhangyang "Atlas" Wang
AlphaPruning: Using Heavy-Tailed Self Regularization Theory for Improved Layer-wise Pruning of Large Language Models
Haiquan Lu · Yefan Zhou · Shiwei Liu · Zhangyang "Atlas" Wang · Michael Mahoney · Yaoqing Yang
Frustratingly Easy Test-Time Adaptation of Vision-Language Models
Matteo Farina · Gianni Franchi · Giovanni Iacca · Massimiliano Mancini · Elisa Ricci
MaskLLM: Learnable Semi-Structured Sparsity for Large Language Models
Gongfan Fang · Hongxu Yin · Saurav Muralidharan · Greg Heinrich · Jeff Pool · Jan Kautz · Pavlo Molchanov · Xinchao Wang
SIRIUS: Contextual Sparisty with Correction for Efficient LLMs
Yang Zhou · Zhuoming Chen · Zhaozhao Xu · Victoria Lin · Beidi Chen
Reducing Transformer Key-Value Cache Size with Cross-Layer Attention
William Brandon · Mayank Mishra · Aniruddha Nrusimha · Rameswar Panda · Jonathan Ragan-Kelley
Search for Efficient Large Language Models
Xuan Shen · Pu Zhao · Yifan Gong · Zhenglun Kong · Zheng Zhan · Yushu Wu · Ming Lin · Chao Wu · Xue Lin · Yanzhi Wang
Efficient Prompt Optimization Through the Lens of Best Arm Identification
Chengshuai Shi · Kun Yang · Zihan Chen · Jundong Li · Jing Yang · Cong Shen
Language Models as Zero-shot Lossless Gradient Compressors: Towards General Neural Parameter Prior Models
Hui-Po Wang · Mario Fritz
Loki: Low-rank Keys for Efficient Sparse Attention
Prajwal Singhania · Siddharth Singh · Shwai He · Soheil Feizi · Abhinav Bhatele
Sparse maximal update parameterization: A holistic approach to sparse training dynamics
Nolan Dey · Shane Bergsma · Joel Hestness
LibAMM: Empirical Insights into Approximate Computing for Accelerating Matrix Multiplication
Xianzhi Zeng · Wenchao Jiang · SHUHAO ZHANG
Reference Trustable Decoding: A Training-Free Augmentation Paradigm for Large Language Models
Luohe Shi · Yao Yao · Zuchao Li · Lefei Zhang · Hai Zhao
\$\textit{Read-ME}\$: Refactorizing LLMs as Router-Decoupled Mixture of Experts with System Co-Design
Ruisi Cai · Yeonju Ro · Geon-Woo Kim · Peihao Wang · Babak Ehteshami Bejnordi · Aditya Akella · Zhangyang "Atlas" Wang
MetaLA: Unified Optimal Linear Approximation to Softmax Attention Map
YUHONG CHOU · Man Yao · Kexin Wang · Yuqi Pan · Rui-Jie Zhu · Jibin Wu · Yiran Zhong · Yu Qiao · Bo Xu · Guoqi Li
Kangaroo: Lossless Self-Speculative Decoding for Accelerating LLMs via Double Early Exiting
Fangcheng Liu · Yehui Tang · Zhenhua Liu · Yunsheng Ni · Duyu Tang · Kai Han · Yunhe Wang
MIInference 1.0: Accelerating Pre-filling for Long-Context LLMs via Dynamic Sparse Attention
Huiqiang Jiang · Yucheng LI · Chengruidong Zhang · Qianhui Wu · Xufang Luo · Surin Ahn · Zhenhua Han · Amir Abdi · Dongsheng Li · Chin-Yew Lin · Yuqing Yang · Lili Qiu
The Mamba in the Llama: Distilling and Accelerating Hybrid Models
Junxiong Wang · Daniele Paliotta · Avner May · Alexander Rush · Tri Dao
Self-Healing Machine Learning: A Framework for Autonomous Adaptation in Real-World Environments
Paulius Rauba · Nabeel Seedat · Krzysztof Kacprzyk · Mihaela van der Schaar
Delta-CoMe: Training-Free Delta-Compression with Mixed-Precision for Large Language Models
Bowen Ping · Shuo Wang · Hanqing Wang · Xu Han · Yuzhuang Xu · Yukun Yan · Yun Chen · Baobao Chang · Zhiyuan Liu · Maosong Sun
OneBit: Towards Extremely Low-bit Large Language Models
Yuzhuang Xu · Xu Han · Zonghan Yang · Shuo Wang · Qingfu Zhu · Zhiyuan Liu · Weidong Liu · Wanxiang Che
Parallelizing Linear Transformers with the Delta Rule over Sequence Length
Songlin Yang · Bailin Wang · Yu Zhang · Yikang Shen · Yoon Kim
ALPS: Improved Optimization for Highly Sparse One-Shot Pruning for Large Language Models
Xiang Meng · Kayhan Behdin · Haoyue Wang · Rahul Mazumder
MoEUT: Mixture-of-Experts Universal Transformers
Róbert Csordás · Kazuki Irie · Jürgen Schmidhuber · Christopher Potts · Christopher D Manning
LISA: Layerwise Importance Sampling for Memory-Efficient Large Language Model Fine-Tuning
Rui Pan · Xiang Liu · SHIZHE DIAO · Renjie Pi · Jipeng Zhang · Chi Han · Tong Zhang
You Only Cache Once: Decoder-Decoder Architectures for Language Models
Yutao Sun · Li Dong · Yi Zhu · Shaohan Huang · Wenhui Wang · Shuming Ma · Quanlu Zhang · Jianyong Wang · Furu Wei
NoMAD-Attention: Efficient LLM Inference on CPUs Through Multiply-add-free Attention
Tianyi Zhang · Jonah Yi · Bowen Yao · Zhaozhao Xu · Anshumali Shrivastava
MATES: Model-Aware Data Selection for Efficient Pretraining with Data Influence Models
Zichun Yu · Spandan Das · Chenyan Xiong

Why Do We Need Weight Decay in Modern Deep Learning?
Francesco D'Angelo · Maksym Andriushchenko · Aditya Vardhan Varre · Nicolas Flammarion
Adversarial Moment-Matching Distillation of Large Language Models
Chen Jia
Efficient Sketches for Training Data Attribution and Studying the Loss Landscape
Andrea Schioppa
FlashAttention-3: Fast and Accurate Attention with Asynchrony and Low-precision
Jay Shah · Ganesh Bikshandi · Ying Zhang · Vijay Thakkar · Pradeep Ramani · Tri Dao
PaDeLLM-NER: Parallel Decoding in Large Language Models for Named Entity Recognition
Jinghui Lu · Yanjie Wang · Ziwei Yang · Xuejing Liu · Brian Mac Namee · Can Huang
LSH-MoE: Communication-efficient MoE Training via Locality-Sensitive Hashing
Xiaonan Nie · Liu Qibin · Fangcheng Fu · Shenhan Zhu · Xupeng Miao · Xiaoyang Li · Yang Zhang · Shouda Liu · Bin CUI
Block Transformer: Global-to-Local Language Modeling for Fast Inference
Namgyu Ho · Sangmin Bae · Taehyeon Kim · Hyunjik Jo · Yireun Kim · Tal Schuster · Adam Fisch · James Thorne · Se-Young Yun
DHA: Learning Decoupled-Head Attention from Transformer Checkpoints via Adaptive Heads Fusion
Yilong Chen · Linhao Zhang · Junyuan Shang · Zhenyu Zhang · Tingwen Liu · Shuohuan Wang · YU SUN
MatFormer: Nested Transformer for Elastic Inference
Fnu Devvrit · Sneha Kudugunta · Aditya Kusupati · Tim Dettmers · Kaifeng Chen · Inderjit Dhillon · Yulia Tsvetkov · Hanna Hajishirzi · Sham Kakade · Ali Farhadi · Prateek Jain
SlimGPT: Layer-wise Structured Pruning for Large Language Models
Gui Ling · Ziyang Wang · Yuliang Yan · Qingwen Liu
MemoryFormer : Minimize Transformer Computation by Removing Fully-Connected Layers
Ning Ding · Yehui Tang · Haochen Qin · Zhenli Zhou · Chao Xu · Lin Li · Kai Han · Liao Heng · Yunhe Wang
Superposed Decoding: Multiple Generations from a Single Autoregressive Inference Pass
Ethan Shen · Alan Fan · Sarah Pratt · Jae Sung Park · Matthew Wallingford · Sham Kakade · Ari Holtzman · Ranjay Krishna · Ali Farhadi · Aditya Kusupati
Towards Exact Gradient-based Training on Analog In-memory Computing
Zhaoxian Wu · Tayfun Gokmen · Malte Rasch · Tianyi Chen
DuQuant: Distributing Outliers via Dual Transformation Makes Stronger Quantized LLMs
Haokun Lin · Haobo Xu · Yichen WU · Jingzhi Cui · Yingtao Zhang · Linzhan Mou · Linqi Song · Zhenan Sun · Ying Wei
Speculative Decoding with CTC-based Draft Model for LLM Inference Acceleration
Zhuofan Wen · Shangdong Gui · Yang Feng
ZipCache: Accurate and Efficient KV Cache Quantization with Salient Token Identification
Yefei He · Luoming Zhang · Weijia Wu · Jing Liu · Hong Zhou · Bohan Zhuang
\$\text{Trans-LoRA}\$: towards data-free Transferable Parameter Efficient Finetuning
Runqian Wang · Soumya Ghosh · David Cox · Diego Antognini · Aude Oliva · Rogerio Feris · Leonid Karlinsky
MiniCache: KV Cache Compression in Depth Dimension for Large Language Models
Akide Liu · Jing Liu · Zizheng Pan · Yefei He · Reza Haffari · Bohan Zhuang
Fine-tuning and Pretrain
The Importance of Online Data: Understanding Preference Fine-tuning via Coverage
Yuda Song · Gokul Swamy · Aarti Singh · J. Bagnell · Wen Sun
Fine-Tuning Large Vision-Language Models as Decision-Making Agents via Reinforcement Learning
Simon Zhai · Hao Bai · Zipeng Lin · Jiayi Pan · Peter Tong · Yifei Zhou · Alane Suhr · Saining Xie · Yann LeCun · Yi Ma · Sergey Levine
Multi-turn Reinforcement Learning with Preference Human Feedback
Lior Shani · Aviv Rosenberg · Asaf Cassel · Oran Lang · Daniele Calandriello · Avital Zipori · Hila Noga · Orgad Keller · Bilal Piot · Idan Szpektor · Avinatan Hassidim · Yossi Matias · Remi Munos
Policy Learning from Tutorial Books via Understanding, Rehearsing and Introspecting
Xiong-Hui Chen · Ziyang Wang · Yali Du · Shengyi Jiang · Meng Fang · Yang Yu · Jun Wang
Risk-Averse Fine-tuning of Large Language Models
Sapana Chaudhary · Ujwal Dinesha · Dileep Kalathil · Srinivas Shakkottai
Learning Goal-Conditioned Representations for Language Reward Models
Vaskar Nath · Dylan Slack · Jeff Da · Yuntao Ma · Hugh Zhang · Spencer Whitehead · Sean Hendryx
Regularizing Hidden States Enables Learning Generalizable Reward Model for LLMs
Rui Yang · Ruomeng Ding · Yong Lin · Huan Zhang · Tong Zhang
Recursive Introspection: Teaching Language Model Agents How to Self-Improve
Yuxiao Qu · Tianjun Zhang · Naman Garg · Aviral Kumar
Imitating Language via Scalable Inverse Reinforcement Learning

Markus Wulfmeier · Michael Bloesch · Nino Vieillard · Arun Ahuja · Jorg Bornschein · Sandy Huang · Artem Sokolov · Matt Barnes · Guillaume Desjardins · Alex Bewley · Sarah Bechtle · Jost Springenberg · Nikola Momchev · Olivier Bachem · Matthieu Geist · Martin Riedmiller
Would I Lie To You? Inference Time Alignment of Language Models using Direct Preference Heads
Avelina Hadji-Kyriacou · Ognjen Arandjelovic
Embedding-Aligned Language Models
Guy Tennenholtz · Yinlam Chow · Chih-wei Hsu · Lior Shani · Yi Liang · Craig Boutilier
A Critical Evaluation of AI Feedback for Aligning Large Language Models
Archit Sharma · Sedrick Scott Keh · Eric Mitchell · Chelsea Finn · Kushal Arora · Thomas Kollar
Decomposed Prompt Decision Transformer for Efficient Unseen Task Generalization
Hongling Zheng · Li Shen · Yong Luo · Tongliang Liu · Jialie Shen · Dacheng Tao
Group Robust Preference Optimization in Reward-free RLHF
Shyam Sundhar Ramesh · Yifan Hu · Jason Chailamas · Viraj Mehta · Pier Giuseppe Sessa · Haitham Bou Ammar · Ilija Bogunovic
Robust Reinforcement Learning from Corrupted Human Feedback
Alexander Bukharin · Ilgee Hong · Haoming Jiang · Zichong Li · Qingru Zhang · Zixuan Zhang · Tuo Zhao
Adaptive Preference Scaling for Reinforcement Learning with Human Feedback
Ilgee Hong · Zichong Li · Alexander Bukharin · Yixiao Li · Haoming Jiang · Tianbao Yang · Tuo Zhao
REBEL: Reinforcement Learning via Regressing Relative Rewards
Zhaolin Gao · Jonathan Chang · Wenhao Zhan · Owen Oertell · Gokul Swamy · Kianté Brantley · Thorsten Joachims · Drew Bagnell · Jason Lee · Wen Sun
SelectIT: Selective Instruction Tuning for LLMs via Uncertainty-Aware Self-Reflection
Liangxin Liu · Xuebo Liu · Derek Wong · Dongfang Li · Ziyi Wang · Baotian Hu · Min Zhang
Tri-Level Navigator: LLM-Empowered Tri-Level Learning for Time Series OOD Generalization
Chengtao Jian · Kai Yang · Yang Jiao
Learn more, but bother less: parameter efficient continual learning
Fuli Qiao · Mehrdad Mahdavi
From Instance Training to Instruction Learning: Task Adapters Generation from Instructions
Huanxuan Liao · Shizhu He · Yao Xu · Yuanzhe Zhang · Yanchao Hao · Shengping Liu · Kang Liu · Jun Zhao
FAST: A Dual-tier Few-Shot Learning Paradigm for Whole Slide Image Classification
Kexue Fu · xiaoyuan luo · Linhao Qu · Shuo Wang · Ying Xiong · Ilias Maglogiannis · Longxiang Gao · Manning Wang
A teacher-teacher framework for clinical language representation learning
Feiqing Huang · Shenghan Zhang · Sara Sweet · Tianxi Cai
Make Your LLM Fully Utilize the Context
Shengnan An · Zexiong Ma · Zeqi Lin · Nanning Zheng · Jian-Guang Lou · Weizhu Chen
Transformers are Minimax Optimal Nonparametric In-Context Learners
Juno Kim · Tai Nakamaki · Taiji Suzuki
Dual-Personalizing Adapter for Federated Foundation Models
Yiyuan Yang · Guodong Long · Tao Shen · Jing Jiang · Michael Blumenstein
The Representation Landscape of Few-Shot Learning and Fine-Tuning in Large Language Models
Diego Doimo · Alessandro Serra · Alessio Ansuini · Alberto Cazzaniga
LACIE: Listener-Aware Finetuning for Calibration in Large Language Models
Elias Stengel-Eskin · Peter Hase · Mohit Bansal
On Softmax Direct Preference Optimization for Recommendation
Yuxin Chen · Junfei Tan · An Zhang · Zhengyi Yang · Leheng Sheng · Enzhi Zhang · Xiang Wang · Tat-Seng Chua
Model Fusion through Bayesian Optimization in Language Model Fine-Tuning
Chaeyun Jang · Hyungi Lee · Jungtaek Kim · Juho Lee
BitDelta: Your Fine-Tune May Only Be Worth One Bit
James Liu · Guangxuan Xiao · Kai Li · Jason Lee · Song Han · Tri Dao · Tianle Cai
How Do Large Language Models Acquire Factual Knowledge During Pretraining?
Hoyeon Chang · Jinho Park · Seonghyeon Ye · Sohee Yang · Youngkyung Seo · Du-Seong Chang · Minjoon Seo
How does Architecture Influence the Base Capabilities of Pre-trained Language Models? A Case Study Based on FFN-Wider and MoE Transformers
Xin Lu · Yanyan Zhao · Bing Qin · Liangyu Huo · Qing Yang · Dongliang Xu
D-CPT Law: Domain-specific Continual Pre-Training Scaling Law for Large Language Models
Haoran Que · Jiaheng Liu · Ge Zhang · Chenchen Zhang · Xingwei Qu · Yinghao Ma · Feiyu Duan · Zhiqi Bai · Zhiqi · Jiakai Wang · Yuanxing Zhang · Xu Tan · Jie Fu · Jiamang Wang · Lin Qu · Wenbo Su · Bo Zheng
Towards a theory of how the structure of language is acquired by deep neural networks
Francesco Cagnetta · Matthieu Wyart
Unveiling LoRA Intrinsic Ranks via Saliency Analysis
Wenjun Ke · Jiahao Wang · Peng Wang · Jiajun Liu · Dong Nie · Guozheng Li · Yining Li
HydraLoRA: An Asymmetric LoRA Architecture for Efficient Fine-Tuning

Chunlin Tian · Zhan Shi · Zhijiang Guo · Li Li · Cheng-Zhong Xu
Time-Reversal Provides Unsupervised Feedback to LLMs
Yerram Varun · Rahul Madhavan · Sravanti Addepalli · Arun Suggala · Karthikeyan Shanmugam · Prateek Jain
Discovering Preference Optimization Algorithms with and for Large Language Models
Chris Lu · Samuel Holt · Claudio Fanconi · Alex Chan · Jakob Foerster · Mihaela van der Schaar · Robert Lange
Not All Tokens Are What You Need for Pretraining
Zhenghao Lin · Zhibin Gou · Yeyun Gong · Xiao Liu · yelong shen · Ruochen Xu · Chen Lin · Yujiu Yang · Jian Jiao · Nan Duan · Weizhu Chen
Customizing Language Models with Instance-wise LoRA for Sequential Recommendation
Xiaoyu Kong · Jiancan Wu · An Zhang · Leheng Sheng · Hui Lin · Xiang Wang · Xiangnan He
TuneTables: Context Optimization for Scalable Prior-Data Fitted Networks
Benjamin Feuer · Robin Schirrmeister · Valeriia Cherepanova · Chinmay Hegde · Frank Hutter · Micah Goldblum · Niv Cohen · Colin White
PISSA: Principal Singular Values and Singular Vectors Adaptation of Large Language Models
Fanxu Meng · Zhaohui Wang · Muhan Zhang
Latent Paraphrasing: Perturbation on Layers Improves Knowledge Injection in Language Models
Minki Kang · Sung Ju Hwang · Gibbeum Lee · Jaewoong Cho
DDK: Distilling Domain Knowledge for Efficient Large Language Models
Jiaheng Liu · Chenchen Zhang · Jinyang Guo · Yuanxing Zhang · Haoran Que · Ken Deng · Zhiqi Bai zhiqi · Jie Liu · Ge Zhang · Jiakai Wang · Yanan Wu · Congnan Liu · Jiamang Wang · Lin Qu · Wenbo Su · Bo Zheng
KG-FIT: Knowledge Graph Fine-Tuning Upon Open-World Knowledge
Pengcheng Jiang · Lang Cao · Cao (Danica) Xiao · Parminder Bhatia · Jimeng Sun · Jiawei Han
End-to-End Ontology Learning with Large Language Models
Andy Lo · Albert Q. Jiang · Wenda Li · Mateja Jamnik
Once Read is Enough: Domain-specific Pretraining-free Language Models with Cluster-guided Sparse Experts for Long-tail Domain Knowledge
Fang Dong · Mengyi Chen · Jixian Zhou · Yubin Shi · Yixuan Chen · Mingzhi Dong · Yujiang Wang · Dongsheng Li · Xiaochen Yang · Rui Zhu · Robert Dick · Qin Lv · Fan Yang · Tun Lu · Ning Gu · Li Shang
Instruction Tuning With Loss Over Instructions
Zhengxiang Shi · Adam Yang · Bin Wu · Laurence Aitchison · Emine Yilmaz · Aldo Lipani
AlchemistCoder: Harmonizing and Eliciting Code Capability by Hindsight Tuning on Multi-source Data
Zifan Song · Yudong Wang · Wenwei Zhang · Kuikun Liu · Chengqi Lyu · Demin Song · Qipeng Guo · Hang Yan · Dahua Lin · Kai Chen · Cairong Zhao
CorDA: Context-Oriented Decomposition Adaptation of Large Language Models for Task-Aware Parameter-Efficient Fine-tuning
Yibo Yang · Xiaojie Li · Zhongzhu Zhou · Shuaiwen Song · Jianlong Wu · Liqiang Nie · Bernard Ghanem
BLoB: Bayesian Low-Rank Adaptation by Backpropagation for Large Language Models
Yibin Wang · Haizhou Shi · Ligong Han · Dimitris Metaxas · Hao Wang
β-DPO: Direct Preference Optimization with Dynamic β
Junkang Wu · Yuexiang Xie · Zhengyi Yang · Jiancan Wu · Jinyang Gao · Bolin Ding · Xiang Wang · Xiangnan He
Reasons and Solutions for the Decline in Model Performance after Editing
Xiusheng Huang · Jiaxiang Liu · Yequan Wang · Kang Liu
Train-Attention: Meta-Learning Where to Focus in Continual Knowledge Learning
Seo Yeongbin · Dongha Lee · Jinyoung Yeo
Is Programming by Example solved by LLMs?
Wen-Ding Li · Kevin Ellis
Federated Fine-tuning of Large Language Models under Heterogeneous Tasks and Client Resources
Jiamu Bai · Daoyuan Chen · Bingchen Qian · Liuyi Yao · Yaliang Li
Bridging The Gap between Low-rank and Orthogonal Adaptation via Householder Reflection Adaptation
Shen Yuan · Haotian Liu · Hongteng Xu
Unpacking DPO and PPO: Disentangling Best Practices for Learning from Preference Feedback
Hamish Ivison · Yizhong Wang · Jiacheng Liu · Zeqiu Wu · Valentina Pyatkin · Nathan Lambert · Noah Smith · Yejin Choi · Hanna Hajishirzi
UniMTS: Unified Pre-training for Motion Time Series
Xiyuan Zhang · Diyan Teng · Ranak Roy Chowdhury · Shuheng Li · Dezhi Hong · Rajesh Gupta · Jingbo Shang
Pre-trained Large Language Models Use Fourier Features to Compute Addition
Tianyi Zhou · Deqing Fu · Vatsal Sharan · Robin Jia
Group and Shuffle: Efficient Structured Orthogonal Parametrization
Mikhail Gorbunov · Nikolay Yudin · Vera Soboleva · Aibek Alanov · Alexey Naumov · Maxim Rakhuba
Large Scale Transfer Learning for Tabular Data via Language Modeling
Josh Gardner · Juan Perdomo · Ludwig Schmidt

CodeRosetta: Pushing the Boundaries of Unsupervised Code Translation for Parallel Programming
Ali TehraniJamsaz · Arijit Bhattacharjee · Le Chen · Nesreen K. Ahmed · Amir Yazdanbakhsh · Ali Jannesari
Personalized Steering of Large Language Models: Versatile Steering Vectors Through Bi-directional Preference Optimization
Yuanpu Cao · Tianrong Zhang · Bochuan Cao · Ziyi Yin · Lu Lin · Fenglong Ma · Jinghui Chen
Talking Heads: Understanding Inter-Layer Communication in Transformer Language Models
Jack Merullo · Carsten Eickhoff · Ellie Pavlick
Enhancing Domain Adaptation through Prompt Gradient Alignment
Viet Hoang Phan · Tung Lam Tran · Quyen Tran · Trung Le
SHED: Shapley-Based Automated Dataset Refinement for Instruction Fine-Tuning
Yexiao He · Ziyao Wang · Zheyu Shen · Guoheng Sun · Yucong Dai · Yongkai Wu · Hongyi Wang · Ang Li
Understanding Emergent Abilities of Language Models from the Loss Perspective
Zhengxiao Du · Aohan Zeng · Yuxiao Dong · Jie Tang
On Giant's Shoulders: Effortless Weak to Strong by Dynamic Logits Fusion
Chenghao Fan · Zhenyi Lu · Wei Wei · Jie Tian · Xiaoye Qu · Dangyang Chen · Yu Cheng
Towards Calibrated Robust Fine-Tuning of Vision-Language Models
Changdae Oh · Hyesu Lim · Mijoo Kim · Dongyoon Han · Sangdoo Yun · Jaegul Choo · Alexander Hauptmann · Zhi-Qi Cheng · Kyungwoo Song
Safe LoRA: The Silver Lining of Reducing Safety Risks when Finetuning Large Language Models
Chia-Yi Hsu · Yu-Lin Tsai · Chih-Hsun Lin · Pin-Yu Chen · Chia-Mu Yu · Chun-Ying Huang
Finding Transformer Circuits With Edge Pruning
Adithya Bhaskar · Alexander Wettig · Dan Friedman · Danqi Chen
TSDS: Data Selection for Task-Specific Model Finetuning
Zifan Liu · Amin Karbasi · Theodoros Rekatsinas
Vision-Language Models are Strong Noisy Label Detectors
Tong Wei · Hao-Tian Li · Chunshu Li · Jiang-Xin Shi · Yu-Feng Li · Min-Ling Zhang
PediatricsGPT: Large Language Models as Chinese Medical Assistants for Pediatric Applications
Dingkang Yang · Jinjie Wei · Dongling Xiao · Shunli Wang · Tong Wu · Gang Li · Mingcheng Li · Shuaibing Wang · Jiawei Chen · Yue Jiang · Qingyao Xu · Ke Li · Peng Zhai · Lihua Zhang
Implicit Optimization Bias of Next-token Prediction in Linear Models
Christos Thrampoulidis
Bandits with Preference Feedback: A Stackelberg Game Perspective
Barna Pásztor · Parnian Kassraie · Andreas Krause
Aggregate-and-Adapt Natural Language Prompts for Downstream Generalization of CLIP
Chen Huang · Skyler Seto · Samira Abnar · David Grangier · Navdeep Jaitly · Joshua Susskind
What Rotary Position Embedding Can Tell Us: Identifying Query and Key Weights Corresponding to Basic Syntactic or High-level Semantic Information
Yiting Chen · Junchi Yan
From News to Forecast: Integrating Event Analysis in LLM-Based Time Series Forecasting with Reflection
Xinlei Wang · Maike Feng · Jing Qiu · Jinjin Gu · Junhua Zhao
VB-LoRA: Extreme Parameter Efficient Fine-Tuning with Vector Banks
Yang Li · Shaobo Han · Shihao Ji
Improving Context-Aware Preference Modeling for Language Models
Silviu Pitis · Ziang Xiao · Nicolas Le Roux · Alessandro Sordoni
CLAP4CLIP: Continual Learning with Probabilistic Finetuning for Vision-Language Models
Saurav Jha · Dong Gong · Lina Yao
ReFT: Representation Finetuning for Language Models
Zhengxuan Wu · Aryaman Arora · Zheng Wang · Atticus Geiger · Dan Jurafsky · Christopher D Manning · Christopher Potts
Deep Bayesian Active Learning for Preference Modeling in Large Language Models
Luckeciano Carvalho Melo · Panagiotis Tigas · Alessandro Abate · Yarin Gal
TAIA: Large Language Models are Out-of-Distribution Data Learners
Shuyang Jiang · Yusheng Liao · Ya Zhang · Yanfeng Wang · Yu Wang
Self-Supervised Alignment with Mutual Information: Learning to Follow Principles without Preference Labels
Jan-Philipp Fraenken · Eric Zelikman · Rafael Rafailov · Kanishk Gandhi · Tobias Gerstenberg · Noah Goodman
Parameter Competition Balancing for Model Merging
Guodong DU · Junlin Lee · Jing Li · Runhua Jiang · Yifei Guo · Shuyang Yu · Hanting Liu · Sim Kuan Goh · Ho-Kin Tang · Daojing He · Min Zhang
Understanding Linear Probing then Fine-tuning Language Models from NTK Perspective
Akiyoshi Tomihari · Issei Sato
SmallToLarge (S2L): Scalable Data Selection for Fine-tuning Large Language Models by Summarizing Training Trajectories of Small Models

Yu Yang · Siddhartha Mishra · Jeffrey Chiang · Baharan Mirzasoleiman
Large Language Models Must Be Taught to Know What They Don't Know
Sanyam Kapoor · Nate Gruver · Manley Roberts · Katie Collins · Arka Pal · Umang Bhatt · Adrian Weller · Samuel Dooley · Micah Goldblum · Andrew Wilson
QuanTA: Efficient High-Rank Fine-Tuning of LLMs with Quantum-Informed Tensor Adaptation
Zhuo Chen · Rumen Dangovski · Charlotte Loh · Owen Dugan · Di Luo · Marin Soljacic
BAdam: A Memory Efficient Full Parameter Optimization Method for Large Language Models
Qijun Luo · Hengxu Yu · Xiao Li
FLoRA: Federated Fine-Tuning Large Language Models with Heterogeneous Low-Rank Adaptations
Ziyao Wang · Zheyu Shen · Yexiao He · Guoheng Sun · Hongyi Wang · Lingjuan Lyu · Ang Li
CoLoR-Filter: Conditional Loss Reduction Filtering for Targeted Language Model Pre-training
David Brandfonbrener · Hanlin Zhang · Andreas Kirsch · Jonathan Richard Schwarz · Sham Kakade
Coevolving with the Other You: Fine-Tuning LLM with Sequential Cooperative Multi-Agent Reinforcement Learning
Hao Ma · Tianyi Hu · Zhiqiang Pu · Liu Boyin · Xiaolin Ai · Yanyan Liang · Min Chen
Quantifying the Gain in Weak-to-Strong Generalization
Moses Charikar · Chirag Pabbaraju · Kirankumar Shiragur
LoFiT: Localized Fine-tuning on LLM Representations
Fangcong Yin · Xi Ye · Greg Durrett
Self-Play Fine-tuning of Diffusion Models for Text-to-image Generation
Huizhuo Yuan · Zixiang Chen · Kaixuan Ji · Quanquan Gu
Scaling Laws for Reward Model Overoptimization in Direct Alignment Algorithms
Rafael Rafailov · Yaswanth Chittepu · Ryan Park · Harshit Sushil Sikchi · Joey Hejna · Brad Knox · Chelsea Finn · Scott Niekum
Interpreting Learned Feedback Patterns in Large Language Models
Luke Marks · Amir Abdullah · Clement Neo · Rauno Arike · David Krueger · Philip Torr · Fazl Barez
Mitigating Reward Overoptimization via Lightweight Uncertainty Estimation
Xiaoying Zhang · Jean-Francois Ton · Wei Shen · Hongning Wang · Yang Liu
Aligning Diffusion Behaviors with Q-functions for Efficient Continuous Control
Huayu Chen · Kaiwen Zheng · Hang Su · Jun Zhu
Interpretability
Do LLMs Build World Representations? Probing Through the Lens of State Abstraction
Zichao Li · Yanshuai Cao · Jackie CK Cheung
Language Models as Hierarchy Encoders
Yuan He · Moy Yuan · Jiaoyan Chen · Ian Horrocks
Instruction Embedding: Latent Representations of Instructions Towards Task Identification
Yiwei Li · Jiayi Shi · Shaoxiong Feng · Peiwen Yuan · Xinglin Wang · Boyuan Pan · Heda Wang · Yao Hu · Prof. Kan
LG-CAV: Train Any Concept Activation Vector with Language Guidance
Qihan Huang · Jie Song · Mengqi Xue · Haofei Zhang · Bingde Hu · Huiqiong Wang · Hao Jiang · Xingen Wang · Mingli Song
Adversarial Representation Engineering: A General Model Editing Framework for Large Language Models
Yihao Zhang · Zeming Wei · Jun Sun · Meng Sun
Towards Understanding How Transformers Learn In-context Through a Representation Learning Lens
Ruifeng Ren · Yong Liu
Reproducibility study of “LICO: Explainable Models with Language-Image Consistency”
Luan Fletcher · Robert van der Klis · Martin Sedlacek · Stefan Vasilev · Christos Athanasiadis
ContextCite: Attributing Model Generation to Context
Benjamin Cohen-Wang · Harshay Shah · Kristian Georgiev · Aleksander Madry
Large Language Models-guided Dynamic Adaptation for Temporal Knowledge Graph Reasoning
Jiapu Wang · Sun Kai · LINHAO LUO · Wei Wei · Yongli Hu · Alan Wee-Chung Liew · Shirui Pan · Baocai Yin
Learning to grok: Emergence of in-context learning and skill composition in modular arithmetic tasks
Tianyu He · Darshil Doshi · Aritra Das · Andrey Gromov
Unified Lexical Representation for Interpretable Visual-Language Alignment
Yifan Li · Yikai Wang · Yanwei Fu · Dongyu Ru · Zheng Zhang · Tong He
Large language model validity via enhanced conformal prediction methods
John Cherian · Isaac Gibbs · Emmanuel Candes
shapiq: Shapley Interactions for Machine Learning
Maximilian Muschalik · Hubert Baniecki · Fabian Fumagalli · Patrick Kolpaczki · Barbara Hammer · Eyke Hüllermeier
VLG-CBM: Training Concept Bottleneck Models with Vision-Language Guidance
Divyansh Srivastava · Ge Yan · Lily Weng
Transformers Represent Belief State Geometry in their Residual Stream
Adam Shai · Paul Riechers · Lucas Teixeira · Alexander Oldenziel · Sarah Marzen
Probing Social Bias in Labor Market Text Generation by ChatGPT: A Masked Language Model Approach

Lei Ding · Yang Hu · Nicole Denier · Enze Shi · Junxi Zhang · Qirui Hu · Karen Hughes · Linglong Kong · Bei Jiang
Divergences between Language Models and Human Brains
Yuchen Zhou · Emmy Liu · Graham Neubig · Michael Tarr · Leila Wehbe
Testing Semantic Importance via Betting
Jacopo Teneggi · Jeremias Sulam
Coarse-to-Fine Concept Bottleneck Models
Konstantinos Panousis · Dino Ienco · Diego Marcos
A Concept-Based Explainability Framework for Large Multimodal Models
Jayneel Parekh · Pegah KHAYATAN · Mustafa Shukor · Alasdair Newson · Matthieu Cord
A Theoretical Perspective for Speculative Decoding Algorithm
Ming Yin · Minshuo Chen · Kaixuan Huang · Mengdi Wang
Conjugated Semantic Pool Improves OOD Detection with Pre-trained Vision-Language Models
Mengyuan Chen · Junyu Gao · Changsheng Xu
From Causal to Concept-Based Representation Learning
Goutham Rajendran · Simon Buchholz · Bryon Aragam · Bernhard Schölkopf · Pradeep Ravikumar
Graph-based Unsupervised Disentangled Representation Learning via Multimodal Large Language Models
Baao Xie · Qiuyu Chen · Yunnan Wang · Zequn Zhang · Xin Jin · Wenjun Zeng
Analysing the Generalisation and Reliability of Steering Vectors
Daniel Tan · David Chanin · Aengus Lynch · Brooks Paige · Dimitrios Kanoulas · Adrià Garriga-Alonso · Robert Kirk
Explaining Datasets in Words: Statistical Models with Natural Language Parameters
Ruiqi Zhong · Heng Wang · Dan Klein · Jacob Steinhardt
Linguistic Collapse: Neural Collapse in (Large) Language Models
Robert Wu · Vardan Papyan
Intruding with Words: Towards Understanding Graph Injection Attacks at the Text Level
Runlin Lei · Yuwei Hu · Yuchen Ren · Zhewei Wei
Crafting Interpretable Embeddings for Language Neuroscience by Asking LLMs Questions
Vinamra Benara · Chandan Singh · John Morris · Richard Antonello · Ion Stoica · Alexander Huth · Jianfeng Gao
Confidence Regulation Neurons in Language Models
Alessandro Stolfo · Ben Wu · Wes Gurnee · Yonatan Belinkov · Xingyi Song · Mrinmaya Sachan · Neel Nanda
A Polar coordinate system represents syntax in large language models
Pablo J. Diego Simon · Stéphane d'Ascoli · Emmanuel Chemla · Yair Lakretz · Jean-Remi King
Knowledge Circuits in Pretrained Transformers
Yunzhi Yao · Ningyu Zhang · Zekun Xi · Mengru Wang · Ziwen Xu · Shumin Deng · Huajun Chen
Context-Aware Testing: A New Paradigm for Model Testing with Large Language Models
Paulius Rauba · Nabeel Seedat · Max Ruiz Luyten · Mihaela van der Schaar
IPO: Interpretable Prompt Optimization for Vision-Language Models
Yingjun Du · Wenfang Sun · Cees Snoek
Transformers need glasses! Information over-squashing in language tasks
Federico Barbero · Andrea Banino · Steven Kapturowski · Dharshan Kumaran · João Madeira Araújo · Oleksandr Vitvitskyi · Razvan Pascanu · Petar Veličković
Language Model as Visual Explainer
Xingyi Yang · Xinchao Wang
Transcoders find interpretable LLM feature circuits
Jacob Dunefsky · Philippe Chlenski · Neel Nanda
Hypothesis Testing the Circuit Hypothesis in LLMs
Claudia Shi · Nicolas Beltran Velez · Achille Nazaret · Carolina Zheng · Adrià Garriga-Alonso · Andrew Jesson · Maggie Makar · David Blei
A Phase Transition between Positional and Semantic Learning in a Solvable Model of Dot-Product Attention
Hugo Cui · Freya Behrens · Florent Krzakala · Lenka Zdeborová
MambaLRP: Explaining Selective State Space Sequence Models
Farnoush Rezaei Jafari · Grégoire Montavon · Klaus-Robert Müller · Oliver Eberle
Selective Attention: Enhancing Transformer through Principled Context Control
Xuechen Zhang · Xiangyu Chang · Mingchen Li · Amit Roy-Chowdhury · Jiasi Chen · Samet Oymak
Understanding Transformers via N-Gram Statistics
Timothy Nguyen
Measuring Progress in Dictionary Learning for Language Model Interpretability with Board Game Models
Adam Karvonen · Benjamin Wright · Can Rager · Rico Angell · Jannik Brinkmann · Logan Smith · Claudio Mayrink Verdun · David Bau · Samuel Marks
Towards Explainable Evaluation Metrics for Machine Translation
Christoph Leiter · Piyawat Lertvittayakumjorn · Marina Fomicheva · Wei Zhao · Yang Gao · Steffen Eger
Free Lunch in Pathology Foundation Model: Task-specific Model Adaptation with Concept-Guided Feature Enhancement

Yanyan Huang · Weiqin Zhao · Yihang Chen · Yu Fu · Lequan Yu
DETAIL: Task DEMonsTration Attribution for Interpretable In-context Learning
Zijian Zhou · Xiaoqiang Lin · Xinyi Xu · Alok Prakash · Daniela Rus · Bryan Kian Hsiang Low
Abrupt Learning in Transformers: A Case Study on Matrix Completion
Pulkit Gopalani · Ekdeep S Lubana · Wei Hu
Unveiling Induction Heads: Provable Training Dynamics and Feature Learning in Transformers
Siyu Chen · Heejune Sheen · Tianhao Wang · Zhuoran Yang
LLM agent
DigiRL: Training In-The-Wild Device-Control Agents with Autonomous Reinforcement Learning
Hao Bai · Yifei Zhou · Jiayi Pan · Mert Cemri · Alane Suhr · Sergey Levine · Aviral Kumar
RL-GPT: Integrating Reinforcement Learning and Code-as-policy
Shaoteng Liu · Haoqi Yuan · Minda Hu · Yanwei Li · Yukang Chen · Shu Liu · Zongqing Lu · Jiaya Jia
KALM: Knowledgeable Agents by Offline Reinforcement Learning from Large Language Model Rollouts
Jing-Cheng Pang · Si-Hang Yang · Kaiyuan Li · Jiayi Zhang · Xiong-Hui Chen · Nan Tang · Yang Yu
Reinforcing LLM Agents via Policy Optimization with Action Decomposition
Muning Wen · Ziyu Wan · Jun Wang · Weinan Zhang · Ying Wen
Language Grounded Multi-agent Reinforcement Learning with Human-interpretable Communication
Huao Li · Hossein Nourkhiz Mahjoub · Behdad Chalaki · Vaishnav Tadiparthi · Kwonjoon Lee · Ehsan Moradi Pari · Charles Lewis · Katia Sycara
StrategyLLM: Large Language Models as Strategy Generators, Executors, Optimizers, and Evaluators for Problem Solving
Chang Gao · Haiyun Jiang · Deng Cai · Shuming Shi · Wai Lam
MAGIS: LLM-Based Multi-Agent Framework for GitHub Issue Resolution
Wei Tao · Yucheng Zhou · Yanlin Wang · Wenqiang Zhang · Hongyu Zhang · Yu Cheng
Gorilla: Large Language Model Connected with Massive APIs
Shishir G Patil · Tianjun Zhang · Xin Wang · Joseph Gonzalez
Cooperation, Competition, and Maliciousness: LLM-Stakeholders Interactive Negotiation
Sahar Abdelnabi · Amr Gomaa · Sarath Sivaprasad · Lea Schönherr · Mario Fritz
Automatically Learning Hybrid Digital Twins of Dynamical Systems
Samuel Holt · Tennison Liu · Mihaela van der Schaar
A Decision-Language Model (DLM) for Dynamic Restless Multi-Armed Bandit Tasks in Public Health
Nikhil Behari · Edwin Zhang · YUNFAN ZHAO · Aparna Taneja · Dheeraj Nagaraj · Milind Tambe
Verified Code Transpilation with LLMs
Sahil Bhatia · Jie Qiu · Niranjana Hasabnis · Sanjit Seshia · Alvin Cheung
Cooperate or Collapse: Emergence of Sustainable Cooperation in a Society of LLM Agents
Giorgio Piatti · Zhijing Jin · Max Kleiman-Weiner · Bernhard Schölkopf · Mrinmaya Sachan · Rada Mihalcea
RoleAgent: Building, Interacting, and Benchmarking High-quality Role-Playing Agents from Scripts
Jiaheng Liu · Zehao Ni · Haoran Que · Sun · Noah Wang · Jian Yang · JiakaiWang · Hongcheng Guo · Z.Y. Peng · Ge Zhang · Jiayi Tian · Xingyuan Bu · Ke Xu · Wenge Rong · Junran Peng · ZHAO-XIANG ZHANG
LLMDFA: Analyzing Dataflow in Code with Large Language Models
Chengpeng Wang · Wuqi Zhang · Zian Su · Xiangzhe Xu · Xiaoheng Xie · Xiangyu Zhang
AutoTimes: Autoregressive Time Series Forecasters via Large Language Models
Yong Liu · Guo Qin · Xiangdong Huang · Jianmin Wang · Mingsheng Long
Can Graph Learning Improve Planning in LLM-based Agents?
Xixi Wu · Yifei Shen · Caihua Shan · Kaitao Song · Siwei Wang · Bohang Zhang · Jiarui Feng · Hong Cheng · Wei Chen · Yun Xiong · Dongsheng Li
On scalable oversight with weak LLMs judging strong LLMs
Zachary Kenton · Noah Siegel · Janos Kramar · Jonah Brown-Cohen · Samuel Albanie · Jannis Bulian · Rishabh Agarwal · David Lindner · Yunhao Tang · Noah Goodman · Rohin Shah
Constrained Human-AI Cooperation: An Inclusive Embodied Social Intelligence Challenge
Weihua Du · Qiushi Lyu · Jiaming Shan · Zhenting Qi · Hongxin Zhang · Sunli Chen · Andi Peng · Tianmin Shu · Kwonjoon Lee · Behzad Dariush · Chuang Gan
VLM Agents Generate Their Own Memories: Distilling Experience into Embodied Programs of Thought
Gabriel Sarch · Lawrence Jang · Michael Tarr · William Cohen · Kenneth Marino · Katerina Fragkiadaki
Agent Planning with World Knowledge Model
Shuofei Qiao · Runnan Fang · Ningyu Zhang · Yuqi Zhu · Xiang Chen · Shumin Deng · Yong Jiang · Pengjun Xie · Fei Huang · Huajun Chen
ChatCam: Empowering Camera Control through Conversational AI
Xinhang Liu · Yu-Wing Tai · Chi-Keung Tang
AMOR: A Recipe for Building Adaptable Modular Knowledge Agents Through Process Feedback
Jian Guan · Wei Wu · zujie wen · Peng Xu · Hongning Wang · Minlie Huang

Chain of Agents: Large Language Models Collaborating on Long-Context Tasks
Yusen Zhang · Ruoxi Sun · Yanfei Chen · Tomas Pfister · Rui Zhang · Sercan Arik
AutoManual: Generating Instruction Manuals by LLM Agents via Interactive Environmental Learning
Minghao Chen · Yihang Li · Yanting Yang · Shiyu Yu · Binbin Lin · Xiaofei He
AvaTaR: Optimizing LLM Agents for Tool Usage via Contrastive Reasoning
Shirley Wu · Shiyu Zhao · Qian Huang · Kexin Huang · Michihiro Yasunaga · Kaidi Cao · Vassilis Ioannidis · Karthik Subbian · Jure Leskovec · James Zou
IQA-EVAL: Automatic Evaluation of Human-Model Interactive Question Answering
Ruosen Li · Ruochen Li · Barry Wang · Xinya Du
Aligning LLM Agents by Learning Latent Preference from User Edits
Ge Gao · Alexey Taymanov · Eduardo Salinas · Paul Mineiro · Dipendra Misra
Long-Horizon Planning for Multi-Agent Robots in Partially Observable Environments
Sid Nayak · Adelmo Morrison Orozco · Marina Have · Jackson Zhang · Vittal Thirumalai · Darren Chen · Aditya Kapoor · Eric Robinson · Karthik Gopalakrishnan · James Harrison · Anuj Mahajan · Brian Ichter · Hamsa Balakrishnan
Star-Agents: Automatic Data Optimization with LLM Agents for Instruction Tuning
Hang Zhou · Yehui Tang · Haochen Qin · Yujie Yang · Renren Jin · Deyi Xiong · Kai Han · Yunhe Wang
Synthetic Programming Elicitation for Text-to-Code in Very Low-Resource Programming and Formal Languages
Federico Mora · Justin Wong · Haley Lepe · Sahil Bhatia · Karim Elmaaroufi · George Varghese · Joseph Gonzalez · Elizabeth Polgreen · Sanjit Seshia
Mobile-Agent-v2: Mobile Device Operation Assistant with Effective Navigation via Multi-Agent Collaboration
Junyang Wang · Haiyang Xu · Haitao Jia · Xi Zhang · Ming Yan · Weizhou Shen · Ji Zhang · Fei Huang · Jitao Sang
UrbanKGent: A Unified Large Language Model Agent Framework for Urban Knowledge Graph Construction
Yansong Ning · Hao Liu
Policy Improvement using Language Feedback Models
Victor Zhong · Dipendra Misra · Xingdi Yuan · Marc-Alexandre Côté
ReEvo: Large Language Models as Hyper-Heuristics with Reflective Evolution
Haoran Ye · Jiarui Wang · Zhiguang Cao · Federico Berto · Chuanbo Hua · HAEYEON KIM · Jinkyoo Park · Guojie Song
HYSYNTH: Context-Free LLM Approximation for Guiding Program Synthesis
Shraddha Barke · Emmanuel Anaya Gonzalez · Saketh Ram Kasibatla · Taylor Berg-Kirkpatrick · Nadia Polikarpova
HYDRA: Model Factorization Framework for Black-Box LLM Personalization
Yuchen Zhuang · Haotian Sun · Yue Yu · Rushi Qiang · Qifan Wang · Chao Zhang · Bo Dai
Smoothie: Label Free Language Model Routing
Neel Guha · Mayee Chen · Trevor Chow · Ishan Khare · Christopher Ré
Large Language Models as Urban Residents: An LLM Agent Framework for Personal Mobility Generation
WANG JIAWEI · Renhe Jiang · Chuang Yang · Zengqing Wu · makoto onizuka · Ryosuke Shibasaki · Noboru Koshizuka · Chuan Xiao
FinCon: A Synthesized LLM Multi-Agent System with Conceptual Verbal Reinforcement for Enhanced Financial Decision Making
Yangyang Yu · Zhiyuan Yao · Haohang Li · Zhiyang Deng · Yuechen Jiang · Yupeng Cao · Zhi Chen · Jordan Suchow · Zhenyu Cui · Rong Liu · Zhaozhao Xu · Denghui Zhang · Koduvayur (Suba) Subbalakshmi · GUOJUN XIONG · Yueru He · Jimin Huang · Dong Li · Qianqian Xie
Code Repair with LLMs gives an Exploration-Exploitation Tradeoff
Hao Tang · Keya Hu · Jin Zhou · Si Cheng Zhong · Wei-Long Zheng · Xujie Si · Kevin Ellis
Leveraging Environment Interaction for Automated PDDL Translation and Planning with Large Language Models
Sadegh Mahdavi · Raquel Aoki · Keyi Tang · Yanshuai Cao
Optimus-1: Hybrid Multimodal Memory Empowered Agents Excel in Long-Horizon Tasks
Zaijing Li · Yuquan Xie · Rui Shao · Gongwei Chen · Dongmei Jiang · Liqiang Nie
Autonomous Agents for Collaborative Task under Information Asymmetry
Wei Liu · Chenxi Wang · YiFei Wang · Zihao Xie · Rennai Qiu · Yufan Dang · Zhuoyun Du · Weize Chen · Cheng Yang · Chen Qian
Richelieu: Self-Evolving LLM-Based Agents for AI Diplomacy
Zhenyu Guan · Xiangyu Kong · Fangwei Zhong · Yizhou Wang
Multi-LLM Debate: Framework, Principals, and Interventions
Andrew Estornell · Yang Liu
CulturePark: Boosting Cross-cultural Understanding in Large Language Models
Cheng Li · Damien Teney · Linyi Yang · Qingsong Wen · Xing Xie · Jindong Wang
Reflective Multi-Agent Collaboration based on Large Language Models
Xiaohe Bo · Zeyu Zhang · Quanyu Dai · Xueyang Feng · Lei Wang · Rui Li · Xu Chen · Ji-Rong Wen
MDAgents: An Adaptive Collaboration of LLMs for Medical Decision-Making
Yubin Kim · Chanwoo Park · Hyewon Jeong · Yik Siu Chan · Xuhai "Orson" Xu · Daniel McDuff · Hyeonhoon Lee · Marzyeh Ghassemi · Cynthia Breazeal · Hae Park
AutoGuide: Automated Generation and Selection of Context-Aware Guidelines for Large Language Model Agents

Yao Fu · Dong-Ki Kim · Jaekyeom Kim · Sungryull Sohn · Lajanugen Logeswaran · Kyunghoon Bae · Honglak Lee
SWE-agent: Agent-Computer Interfaces Enable Automated Software Engineering
John Yang · Carlos Jimenez · Alexander Wettig · Kilian Lieret · Shunyu Yao · Karthik Narasimhan · Ofir Press
Can Large Language Model Agents Simulate Human Trust Behavior?
Chengxing Xie · Canyu Chen · Feiran Jia · Ziyu Ye · Shiyang Lai · Kai Shu · Jindong Gu · Adel Bibi · Ziniu Hu · David Jurgens · James Evans · Philip Torr · Bernard Ghanem · Guohao Li
Learning to Discuss Strategically: A Case Study on One Night Ultimate Werewolf
Xuanfa Jin · Ziyang Wang · Yali Du · Meng Fang · Haifeng Zhang · Jun Wang
LLM Augmentation/RAG
Boosting the Potential of Large Language Models with an Intelligent Information Assistant
Yujia Zhou · Zheng Liu · Zhicheng Dou
Ad Auctions for LLMs via Retrieval Augmented Generation
MohammadTaghi Hajiaghayi · Sébastien Lahaie · Keivan Rezaei · Suho Shin
Scaling Retrieval-Based Language Models with a Trillion-Token Datastore
Rulin Shao · Jacqueline He · Akari Asai · Weijia Shi · Tim Dettmers · Sewon Min · Luke Zettlemoyer · Pang Wei Koh
GraphVis: Boosting LLMs with Visual Knowledge Graph Integration
Yihe Deng · Chenchen Ye · Zijie Huang · Mingyu Derek Ma · Yiwon Kou · Wei Wang
Exploratory Retrieval-Augmented Planning For Continual Embodied Instruction Following
Minjong Yoo · Jinwoo Jang · Wei-Jin Park · Honguk Woo
LLM-ESR: Large Language Models Enhancement for Long-tailed Sequential Recommendation
Qidong Liu · Xian Wu · Yejing Wang · Zijian Zhang · Feng Tian · Yefeng Zheng · Xiangyu Zhao
Mobility-LLM: Learning Visiting Intentions and Travel Preference from Human Mobility Data with Large Language Models
Letian Gong · Yan Lin · Xinyue Zhang · Yiwon Lu · Xuedi Han · Yichen Liu · Shengnan Guo · Youfang Lin · Huaiyu Wan
MSA Generation with Seqs2Seqs Pretraining: Advancing Protein Structure Predictions
LE ZHANG · Jiayang Chen · Tao Shen · Yu Li · Siqi Sun
Entity Alignment with Noisy Annotations from Large Language Models
CHEN SHENGYUAN · Qinggang Zhang · Junnan Dong · Wen Hua · Qing Li · Xiao Huang
SearchLVLMs: A Plug-and-Play Framework for Augmenting Large Vision-Language Models by Searching Up-to-Date Internet Knowledge
Chuanhao Li · Zhen Li · Chenchen Jing · Shuo Liu · Wenqi Shao · Yuwei Wu · Ping Luo · Yu Qiao · Kaipeng Zhang
A Textbook Remedy for Domain Shifts: Knowledge Priors for Medical Image Analysis
Yue Yang · Mona Gandhi · Yufei Wang · Yifan Wu · Michael Yao · Chris Callison-Burch · James Gee · Mark Yatskar
Cloud Object Detector Adaptation by Integrating Different Source Knowledge
Shuaifeng Li · Mao Ye · Lihua Zhou · Nianxin Li · Siying Xiao · Song Tang · Xiatian Zhu
Unified Generative and Discriminative Training for Multi-modal Large Language Models
Wei Chow · Juncheng Li · Qifan Yu · Kaihang Pan · Hao Fei · Zhiqi Ge · Shuaiyang · Siliang Tang · Hanwang Zhang · QIANRU SUN
WISE: Rethinking the Knowledge Memory for Lifelong Model Editing of Large Language Models
Peng Wang · Zexi Li · Ningyu Zhang · Ziwen Xu · Yunzhi Yao · Yong Jiang · Pengjun Xie · Fei Huang · Huajun Chen
TableRAG: Million-Token Table Understanding with Language Models
Si-An Chen · Lesly Miculicich · Julian Eisenschlos · Zifeng Wang · Zilong Wang · Yanfei Chen · YASUHISA FUJII · Hsuan-Tien Lin · Chen-Yu Lee · Tomas Pfister
HippoRAG: Neurobiologically Inspired Long-Term Memory for Large Language Models
Bernal Jimenez Gutierrez · Yiheng Shu · Yu Gu · Michihiro Yasunaga · Yu Su
Cost-efficient Knowledge-based Question Answering with Large Language Models
Junnan Dong · Qinggang Zhang · Chuang Zhou · Hao Chen · Daochen Zha · Xiao Huang
Knowledge-Empowered Dynamic Graph Network for Irregularly Sampled Medical Time Series
Yicheng Luo · Zhen Liu · Linghao Wang · Binqian Wu · Junhao Zheng · Qianli Ma
RouterDC: Query-Based Router by Dual Contrastive Learning for Assembling Large Language Models
Shuhao Chen · Weisen Jiang · Baijiong Lin · James Kwok · Yu Zhang
Improving Gloss-free Sign Language Translation by Reducing Representation Density
Jinhui Ye · Xing Wang · Wenxiang Jiao · Junwei Liang · Hui Xiong
Online Adaptation of Language Models with a Memory of Amortized Contexts
Jihoon Tack · Jaehyung Kim · Eric Mitchell · Jinwoo Shin · Yee Whye Teh · Jonathan Richard Schwarz
Nearest Neighbor Speculative Decoding for LLM Generation and Attribution
Minghan Li · Xilun Chen · Ari Holtzman · Beidi Chen · Jimmy Lin · Scott Yih · Victoria Lin
KnowGPT: Knowledge Graph based Prompting for Large Language Models
Qinggang Zhang · Junnan Dong · Hao Chen · Daochen Zha · Zailiang Yu · Xiao Huang
Approaching Human-Level Forecasting with Language Models
Danny Halawi · Fred Zhang · Chen Yueh-Han · Jacob Steinhardt
SlowFocus: Enhancing Fine-grained Temporal Understanding in Video LLM

Ming Nie · Dan Ding · Chunwei Wang · Yuanfan Guo · Jianhua Han · Hang Xu · Li Zhang
RankRAG: Unifying Context Ranking with Retrieval-Augmented Generation in LLMs
Yue Yu · Wei Ping · Zihan Liu · Boxin Wang · Jiaxuan You · Chao Zhang · Mohammad Shoeybi · Bryan Catanzaro
Meta-Diffu\$B\$: A Contextualized Sequence-to-Sequence Text Diffusion Model with Meta-Exploration
Yun-Yen Chuang · Hung-Min Hsu · Kevin Lin · Chen-Sheng Gu · Ling-Zhen Li · Ray-I Chang · Hung-yi Lee
Generated and Pseudo Content guided Prototype Refinement for Few-shot Point Cloud Segmentation
Lili Wei · Congyan Lang · Ziyi Chen · Tao Wang · Yidong Li · Jun Liu
CausalStock: Deep End-to-end Causal Discovery for News-driven Multi-stock Movement Prediction
Shuqi Li · Yuebo Sun · Yuxin Lin · Xin Gao · Shuo Shang · Rui Yan
G-Retriever: Retrieval-Augmented Generation for Textual Graph Understanding and Question Answering
Xiaoxin He · Yijun Tian · Yifei Sun · Nitesh Chawla · Thomas Laurent · Yann LeCun · Xavier Bresson · Bryan Hooi
MKGL: Mastery of a Three-Word Language
Lingbing Guo · Zhongpu Bo · Zhuo Chen · Yichi Zhang · Jiaoyan Chen · Lan Yarong · Mengshu Sun · Zhiqiang Zhang · Yangyifei Luo · Qian Li · Qiang Zhang · Wen Zhang · Huajun Chen
Construction and Application of Materials Knowledge Graph in Multidisciplinary Materials Science via Large Language Model
Yanpeng Ye · Jie Ren · Shaozhou Wang · Yuwei Wan · Imran Razzak · Bram Hoex · Haofen Wang · Tong Xie · Wenjie Zhang
Self-Retrieval: End-to-End Information Retrieval with One Large Language Model
Qiaoyu Tang · Jiawei Chen · Zhuoqun Li · Bowen Yu · Yaojie Lu · ChengFu · Haiyang Yu · Hongyu Lin · Fei Huang · Ben He · Xianpei Han · Le Sun · Yongbin Li
CultureLLM: Incorporating Cultural Differences into Large Language Models
Cheng Li · Mengzhuo Chen · Jindong Wang · Sunayana Sitaram · Xing Xie
CRAG - Comprehensive RAG Benchmark
Xiao Yang · Kai Sun · Hao Xin · Yushi Sun · Nikita Bhalla · Xiangsen Chen · Sajal Choudhary · Rongze Gui · Ziran Jiang · Ziyu Jiang · Lingkun Kong · Brian Moran · Jiaqi Wang · Yifan Xu · An Yan · Chenyu Yang · Eting Yuan · Hanwen Zha · Nan Tang · Lei Chen · Nicolas Scheffer · Yue Liu · Nirav Shah · Rakesh Wanga · Anuj Kumar · Scott Yih · Xin Dong
xRAG: Extreme Context Compression for Retrieval-augmented Generation with One Token
Xin Cheng · Xun Wang · Xingxing Zhang · Tao Ge · Si-Qing Chen · Furu Wei · Huishuai Zhang · Dongyan Zhao
Textual Training for the Hassle-Free Removal of Unwanted Visual Data: Case Studies on OOD and Hateful Image Detection
Saehyung Lee · Jisoo Mok · Sangha Park · Yongho Shin · Dahuin Jung · Sungroh Yoon
ClashEval: Quantifying the tug-of-war between an LLM's internal prior and external evidence
Kevin Wu · Eric Wu · James Zou
Multimodal LLM
Boosting Text-to-Video Generative Model with MLLMs Feedback
Xun Wu · Shaohan Huang · Guolong Wang · Jing Xiong · Furu Wei
Multimodal Large Language Models Make Text-to-Image Generative Models Align Better
Xun Wu · Shaohan Huang · Guolong Wang · Jing Xiong · Furu Wei
Aligning Vision Models with Human Aesthetics in Retrieval: Benchmarks and Algorithms
Miaosen Zhang · Yixuan Wei · Zhen Xing · Yifei Ma · Zuxuan Wu · Ji Li · Zheng Zhang · Qi Dai · Chong Luo · Xin Geng · Baining Guo
GenRL: Multimodal-foundation world models for generalization in embodied agents
Pietro Mazzaglia · Tim Verbelen · Bart Dhoedt · Aaron Courville · Sai Rajeswar Mudumba
Scene Graph Generation with Role-Playing Large Language Models
Guikun Chen · Jin Li · Wenguan Wang
Discrete Modeling via Boundary Conditional Diffusion Processes
Yuxuan Gu · Xiaocheng Feng · Lei Huang · Yingsheng Wu · Zekun Zhou · Weihong Zhong · kun Zhu · Bing Qin
M\$^3\$GPT: An Advanced Multimodal, Multitask Framework for Motion Comprehension and Generation
Mingshuang Luo · RuiBing Hou · Zhuo Li · Hong Chang · Zimo Liu · Yaowei Wang · Shiguang Shan
Unleashing Region Understanding in Intermediate Layers for MLLM-based Referring Expression Generation
Yaoyuan Liang · Zhuojun Cai · Jian Xu · Guanbo Huang · Yiran Wang · Xiao Liang · Jiahao Liu · Ziran Li · Jingang Wang · Shao-Lun Huang
XMask3D: Cross-modal Mask Reasoning for Open Vocabulary 3D Semantic Segmentation
Ziyi Wang · Yanbo Wang · Xumin Yu · Jie Zhou · Jiwen Lu
ShareGPT4Video: Improving Video Understanding and Generation with Better Captions
Lin Chen · Xilin Wei · Jinsong Li · Xiaoyi Dong · Pan Zhang · Yuhang Zang · Zehui Chen · Haodong Duan · lin bin · Zhenyu Tang · Li Yuan · Dahua Lin · Feng Zhao · Jiaqi Wang
Visual Perception by Large Language Model's Weights
Feipeng Ma · Hongwei Xue · Yizhou Zhou · Guangting Wang · Fengyun Rao · Shilin Yan · Yueyi Zhang · Siying Wu · Mike Zheng Shou · Xiaoyan Sun
SPIQA: A Dataset for Multimodal Question Answering on Scientific Papers

Shraman Pramanick · Rama Chellappa · Subhashini Venugopalan
Simplified and Generalized Masked Diffusion for Discrete Data
Jiaxin Shi · Kehang Han · Zhe Wang · Arnaud Doucet · Michalis Titsias
One Token to Seg Them All: Language Instructed Reasoning Segmentation in Videos
Zeichen Bai · Tong He · Haiyang Mei · Pichao WANG · Ziteng Gao · Joya Chen · liulei · Zheng Zhang · Mike Zheng Shou
What matters when building vision-language models?
Hugo Laurençon · Leo Tronchon · Matthieu Cord · Victor Sanh
Exploiting Descriptive Completeness Prior for Cross Modal Hashing with Incomplete Labels
Haoyang Luo · Zheng Zhang · Yadan Luo
Wings: Learning Multimodal LLMs without Text-only Forgetting
Yi-Kai Zhang · Shiyin Lu · Yang Li · YanQing Ma · Qingguo Chen · Zhao Xu · Weihua Luo · Kaifu Zhang · De-Chuan Zhan · Han-Jia Ye
Empowering Visible-Infrared Person Re-Identification with Large Foundation Models
Zhangyi Hu · Bin Yang · Mang Ye
Is A Picture Worth A Thousand Words? Delving Into Spatial Reasoning for Vision Language Models
Jiayu Wang · Yifei Ming · Zhenmei Shi · Vibhav Vineet · Xin Wang · Sharon Li · Neel Joshi
HAWK: Learning to Understand Open-World Video Anomalies
Jiaqi Tang · Hao LU · RUIZHENG WU · Xiaogang Xu · Ke Ma · Cheng Fang · Bin Guo · Jiangbo Lu · Qifeng Chen · Yingcong Chen
Leveraging Visual Tokens for Extended Text Contexts in Multi-Modal Learning
Alex Jinpeng Wang · Linjie Li · Yiqi Lin · Min Li · Lijuan Wang · Mike Zheng Shou
Neural Residual Diffusion Models for Deep Scalable Vision Generation
Zhiyuan Ma · Liangliang Zhao · Biqing Qi · Bowen Zhou
AWT: Transferring Vision-Language Models via Augmentation, Weighting, and Transportation
Yuhan Zhu · Yuyang Ji · Zhiyu Zhao · Gangshan Wu · Limin Wang
Accelerating Pre-training of Multimodal LLMs via Chain-of-Sight
Ziyuan Huang · Kaixiang Ji · Biao Gong · Zhiwu Qing · Qinglong Zhang · Kecheng Zheng · Jian Wang · Jingdong Chen · Ming Yang
Déjà Vu Memorization in Vision–Language Models
Bargav Jayaraman · Chuan Guo · Kamalika Chaudhuri
No Filter: Cultural and Socioeconomic Diversity in Contrastive Vision-Language Models
Angéline Pouget · Lucas Beyer · Emanuele Bugliarello · Xiao Wang · Andreas Steiner · Xiaohua Zhai · Ibrahim Alabdulmohsin
VMamba: Visual State Space Model
Yue Liu · Yunjie Tian · Yuzhong Zhao · Hongtian Yu · Lingxi Xie · Yaowei Wang · Qixiang Ye · Jianbin Jiao · Yunfan Liu
Chat-Scene: Bridging 3D Scene and Large Language Models with Object Identifiers
Haifeng Huang · Yilun Chen · Zehan Wang · Rongjie Huang · Runsen Xu · Tai WANG · Luping Liu · Xize Cheng · Yang Zhao · Jiangmiao Pang · Zhou Zhao
Dense Connector for MLLMs
Huanjin Yao · Wenhao Wu · Taojiannan Yang · YuXin Song · Mengxi Zhang · Haocheng Feng · Yifan Sun · Zhiheng Li · Wanli Ouyang · Jingdong Wang
Why are Visually-Grounded Language Models Bad at Image Classification?
Yuhui Zhang · Alyssa Unell · Xiaohan Wang · Dhruba Ghosh · Yuchang Su · Ludwig Schmidt · Serena Yeung
Paralinguistics-Aware Speech-Empowered Large Language Models for Natural Conversation
Heeseung Kim · Soonshin Seo · Kyeongseok Jeong · Ohsung Kwon · Soyeon Kim · Jungwhan Kim · Jaehong Lee · Eunwoo Song · Myungwoo Oh · Jung-Woo Ha · Sungroh Yoon · Kang Min Yoo
RaVL: Discovering and Mitigating Spurious Correlations in Fine-Tuned Vision-Language Models
Maya Varma · Jean-Benoit Delbrouck · Zhihong Chen · Akshay Chaudhari · Curtis Langlotz
EZ-HOI: VLM Adaptation via Guided Prompt Learning for Zero-Shot HOI Detection
Qinqian Lei · Bo Wang · Robby Tan
Artemis: Towards Referential Understanding in Complex Videos
Jihao Qiu · Yuan Zhang · Xi Tang · Lingxi Xie · Tianren Ma · Pengyu Yan · DAVID DOERMANN · Qixiang Ye · Yunjie Tian
Image Textualization: An Automatic Framework for Generating Rich and Detailed Image Descriptions
Renjie Pi · Jianshu Zhang · Jipeng Zhang · Rui Pan · Zhekai Chen · Tong Zhang
CLIPCEIL: Domain Generalization through CLIP via Channel rEfinement and Image-text aLignment
Xi Yu · Shinjae Yoo · Yuewei Lin
ChatTracker: Enhancing Visual Tracking Performance via Chatting with Multimodal Large Language Model
Yiming Sun · Fan Yu · Shaoxiang Chen · Yu Zhang · Junwei Huang · Yang Li · Chenhui Li · Changbo Wang
OW-VISCapTor: Abstractors for Open-World Video Instance Segmentation and Captioning
Anwesa Choudhuri · Girish Chowdhary · Alex Schwing
Procedure-Aware Surgical Video-language Pretraining with Hierarchical Knowledge Augmentation
Kun Yuan · vinkle srivastav · Nassir Navab · Nicolas Padoy
Few-Shot Adversarial Prompt Learning on Vision-Language Models
Yiwei Zhou · Xiaobo Xia · Zhiwei Lin · Bo Han · Tongliang Liu

Dual Prototype Evolving for Test-Time Generalization of Vision-Language Models
Ce Zhang · Simon Stepputtis · Katia Sycara · Yaqi Xie
DA-Ada: Learning Domain-Aware Adapter for Domain Adaptive Object Detection
Haochen Li · Rui Zhang · Hantao Yao · Xin Zhang · Yifan Hao · Xinkai Song · Xiaqing Li · Yongwei Zhao · Yunji Chen · Ling Li
TabPedia: Towards Comprehensive Visual Table Understanding with Concept Synergy
Weichao Zhao · Hao Feng · Qi Liu · Jingqun Tang · Binghong Wu · Lei Liao · Shu Wei · Yongjie Ye · Hao Liu · Wengang Zhou · Houqiang Li · Can Huang
Boosting Weakly Supervised Referring Image Segmentation via Progressive Comprehension
Zaiquan Yang · Yuhao LIU · Jiaying Lin · Gerhard Hancke · Rynson Lau
Implicit Multimodal Alignment: On the Generalization of Frozen LLMs to Multimodal Inputs
Mustafa Shukor · Matthieu Cord
Unleashing the Potential of the Diffusion Model in Few-shot Semantic Segmentation
Muzhi Zhu · Yang Liu · Zekai Luo · Chenchen Jing · Hao Chen · Guangkai Xu · Xinlong Wang · Chunhua Shen
Visual Anchors Are Strong Information Aggregators For Multimodal Large Language Model
Haogeng Liu · Quanzeng You · Xiaotian Han · Yongfei Liu · Huaibo Huang · Ran He · Hongxia Yang
Vitron: A Unified Pixel-level Vision LLM for Understanding, Generating, Segmenting, Editing
Hao Fei · Shengqiong Wu · Hanwang Zhang · Tat-Seng Chua · Shuicheng Yan
CigTime: Corrective Instruction Generation Through Inverse Motion Editing
Qihang Fang · Chengcheng Tang · Bugra Tekin · Yanchao Yang
Streaming Long Video Understanding with Large Language Models
Rui Qian · Xiaoyi Dong · Pan Zhang · Yuhang Zang · Shuangrui Ding · Dahua Lin · Jiaqi Wang
VisionLLM v2: An End-to-End Generalist Multimodal Large Language Model for Hundreds of Vision-Language Tasks
Giannan Wu · Muyan Zhong · Sen Xing · Zeqiang Lai · Zhaoyang Liu · Zhe Chen · Wenhai Wang · Xizhou Zhu · Lewei Lu · Tong Lu · Ping Luo · Yu Qiao · Jifeng Dai
InternLM-XComposer2-4KHD: A Pioneering Large Vision-Language Model Handling Resolutions from 336 Pixels to 4K HD
Xiaoyi Dong · Pan Zhang · Yuhang Zang · Yuhang Cao · Bin Wang · Linke Ouyang · Songyang Zhang · Haodong Duan · Wenwei Zhang · Yining Li · Hang Yan · Yang Gao · Zhe Chen · xinyue zhang · Wei Li · Li Jingwen · Wenhai Wang · Kai Chen · Conghui He · Xingcheng ZHANG · Jifeng Dai · Yu Qiao · Dahua Lin · Jiaqi Wang
Towards Open-Vocabulary Semantic Segmentation Without Semantic Labels
Heeseong Shin · Chaehyun Kim · Sunghwan Hong · Seokju Cho · Anurag Arnab · Paul Hongsuck Seo · Seungryong Kim
LLaMo: Large Language Model-based Molecular Graph Assistant
Jinyoung Park · Minseong Bae · Dohwan Ko · Hyunwoo Kim
What Factors Affect Multi-Modal In-Context Learning? An In-Depth Exploration
Libo Qin · Qiguang Chen · Hao Fei · Zhi Chen · Min Li · Wanxiang Che
Understanding the Limits of Vision Language Models Through the Lens of the Binding Problem
Declan Campbell · Sunayana Rane · Tyler Giallanza · Camillo Nicolò De Sabbata · Kia Ghods · Amogh Joshi · Alexander Ku · Steven Frankland · Tom Griffiths · Jonathan D Cohen · Taylor Webb
Relationship Prompt Learning is Enough for Open-Vocabulary Semantic Segmentation
Li Jiahao · Yang Lu · Yuan Xie · Yanyun Qu
MeshXL: Neural Coordinate Field for Generative 3D Foundation Models
Sijin Chen · Xin Chen · Anqi Pang · Xianfang Zeng · Wei Cheng · Yijun Fu · Fukun Yin · Billzb Wang · Jingyi Yu · Gang Yu · BIN FU · Tao Chen
Training an Open-Vocabulary Monocular 3D Detection Model without 3D Data
Rui Huang · Henry Zheng · Yan Wang · Zhuofan Xia · Marco Pavone · Gao Huang
InterDreamer: Zero-Shot Text to 3D Dynamic Human-Object Interaction
Sirui Xu · ziyin wang · Yu-Xiong Wang · Liangyan Gui
Lever LM: Configuring In-Context Sequence to Lever Large Vision Language Models
Xu Yang · Yingzhe Peng · Haoxuan Ma · Shuo Xu · Chi Zhang · Yucheng Han · Hanwang Zhang
FlexCap: Describe Anything in Images in Controllable Detail
Debidatta Dwibedi · Vidhi Jain · Jonathan Tompson · Andrew Zisserman · Yusuf Aytar
Self-Calibrated Tuning of Vision-Language Models for Out-of-Distribution Detection
Geng Yu · Jianing Zhu · Jiangchao Yao · Bo Han
Hierarchical Visual Feature Aggregation for OCR-Free Document Understanding
JaeYoo Park · Jin Young Choi · Jeonghyung Park · Bohyung Han
Advancing Cross-domain Discriminability in Continual Learning of Vision-Language Models
Yicheng Xu · Yuxin Chen · Jiahao Nie · Yusong Wang · HUIPING ZHUANG · Manabu Okumura
Single Image Unlearning: Efficient Machine Unlearning in Multimodal Large Language Models
Jiaqi Li · Qianshan Wei · Chuanyi Zhang · Guilin Qi · Miaozen Du · Yongrui Chen · Sheng Bi · Fan Liu
MoME: Mixture of Multimodal Experts for Generalist Multimodal Large Language Models
Leyang Shen · Gongwei Chen · Rui Shao · Weili Guan · Liqiang Nie
Prism: A Framework for Decoupling and Assessing the Capabilities of VLMs

Yuxuan Qiao · Haodong Duan · Xinyu Fang · Junming Yang · Lin Chen · Songyang Zhang · Jiaqi Wang · Dahua Lin · Kai Chen
Neuro-Vision to Language: Enhancing Brain Recording-based Visual Reconstruction and Language Interaction
Guobin Shen · Dongcheng Zhao · Xiang He · Linghao Feng · Yiting Dong · Jihang Wang · Qian Zhang · Yi Zeng
Multi-modal Transfer Learning between Biological Foundation Models
Juan Jose Garau-Luis · Patrick Bordes · Liam Gonzalez · Maša Roller · Bernardo de Almeida · Christopher Blum · Lorenz Hexemer · Stefan Laurent · Maren Lang · Thomas PIERROT · Guillaume Richard
CogVLM: Visual Expert for Pretrained Language Models
Weihan Wang · Qingsong Lv · Wenmeng Yu · Wenyi Hong · Ji Qi · Yan Wang · Junhui Ji · Zhuoyi Yang · Lei Zhao · Song XiXuan · Jiazheng Xu · Keqin Chen · Bin Xu · Juanzi Li · Yuxiao Dong · Ming Ding · Jie Tang
DiPEX: Dispersing Prompt Expansion for Class-Agnostic Object Detection
Jia S Lim · Zhuoxiao Chen · Zhi Chen · Mahsa Baktashmotlagh · Xin Yu · Zi Huang · Yadan Luo
VLMimic: Vision Language Models are Visual Imitation Learner for Fine-grained Actions
Guangyan Chen · Meiling Wang · Te Cui · Yao Mu · Haoyang Lu · Tianxing Zhou · Zicai Peng · Mengxiao Hu · Haizhou Li · Li Yuan · Yi Yang · Yufeng Yue
LOVA3: Learning to Visual Question Answering, Asking and Assessment
Henry Hengyuan Zhao · Pan Zhou · Difei Gao · Zechen Bai · Mike Zheng Shou
InterControl: Zero-shot Human Interaction Generation by Controlling Every Joint
Zhenzhi Wang · Jingbo Wang · Yixuan Li · Dahua Lin · Bo Dai
How do Large Language Models Handle Multilingualism?
Yiran Zhao · Wenxuan Zhang · Guizhen Chen · Kenji Kawaguchi · Lidong Bing
PromptFix: You Prompt and We Fix the Photo
yongsheng yu · Ziyun Zeng · Hang Hua · Jianlong Fu · Jiebo Luo
Unveiling Encoder-Free Vision-Language Models
Haiwen Diao · Yufeng Cui · Xiaotong Li · Yueze Wang · Huchuan Lu · Xinlong Wang
DeTikZify: Synthesizing Graphics Programs for Scientific Figures and Sketches with TikZ
Jonas Belouadi · Simone Ponzetto · Steffen Eger
LLaNA: Large Language and NeRF Assistant
Andrea Amaduzzi · Pierluigi Zama Ramirez · Giuseppe Lisanti · Samuele Salti · Luigi Di Stefano
VLKEB: A Large Vision-Language Model Knowledge Editing Benchmark
Han Huang · Haitian Zhong · Tao Yu · Qiang Liu · Shu Wu · Liang Wang · Tieniu Tan
Lumen: Unleashing Versatile Vision-Centric Capabilities of Large Multimodal Models
Yang Jiao · Shaoxiang Chen · Zequn Jie · Jingjing Chen · Lin Ma · Yu-Gang Jiang
Understanding Information Storage and Transfer in Multi-Modal Large Language Models
Samyadeep Basu · Martin Grayson · Cecily Morrison · Besmira Nushi · Soheil Feizi · Daniela Massiceti
CLIPLoss and Norm-Based Data Selection Methods for Multimodal Contrastive Learning
Yiping Wang · Yifang Chen · Wendan Yan · Alex Fang · Wenjing Zhou · Kevin Jamieson · Simon Du
Towards Neuron Attributions in Multi-Modal Large Language Models
Junfeng Fang · Zac Bi · Ruipeng Wang · Houcheng Jiang · Yuan Gao · Kun Wang · An Zhang · Jie Shi · Xiang Wang · Tat-Seng Chua
Efficient Large Multi-modal Models via Visual Context Compression
Jieneng Chen · Luoxin Ye · Ju He · Zhaoyang Wang · Daniel Khashabi · Alan Yuille
OmniJARVIS: Unified Vision-Language-Action Tokenization Enables Open-World Instruction Following Agents
Zihao Wang · Shaofei Cai · Zhancun Mu · Haowei Lin · Ceyao Zhang · Xuejie Liu · Qing Li · Anji Liu · Xiaojian (Shawn) Ma · Yitao Liang
Training-Free Open-Ended Object Detection and Segmentation via Attention as Prompts
Zhiwei Lin · Yongtao Wang · Zhi Tang
Easy Regional Contrastive Learning of Expressive Fashion Representations
Daiqing Qi · Handong Zhao · Sheng Li
UniAudio 1.5: Large Language Model-Driven Audio Codec is A Few-Shot Audio Task Learner
Dongchao Yang · Haohan Guo · Yuanyuan Wang · Rongjie Huang · Xiang Li · Xu Tan · Xixin Wu · Helen Meng
GITA: Graph to Visual and Textual Integration for Vision-Language Graph Reasoning
Yanbin Wei · Shuai Fu · Weisen Jiang · Zejian Zhang · Zhixiong Zeng · Qi Wu · James Kwok · Yu Zhang
LIVE: Learnable In-Context Vector for Visual Question Answering
Yingzhe Peng · chenduo hao · Xinting Hu · Jiawei Peng · Xin Geng · Xu Yang
Learning 1D Causal Visual Representation with De-focus Attention Networks
Tao Chenxin · Xizhou Zhu · Shiqian Su · Lewei Lu · Changyao Tian · Xuan Luo · Gao Huang · Hongsheng Li · Yu Qiao · Jie Zhou · Jifeng Dai
Seeing the Image: Prioritizing Visual Correlation by Contrastive Alignment
Xin Xiao · Bohong Wu · Jiacong Wang · Chunyuan Li · Zhou Xun · Haoyuan Guo
Hidden in Plain Sight: Evaluating Abstract Shape Recognition in Vision-Language Models
Arshia Hemmat · Adam Davies · Tom Lamb · Jianhao Yuan · Philip Torr · Ashkan Khakzar · Francesco Pinto

Boosting Vision-Language Models with Transduction
Maxime Zanella · Benoît Gérin · Ismail Ayed
Emotion-LLaMA: Multimodal Emotion Recognition and Reasoning with Instruction Tuning
Zebang Cheng · Zhi-Qi Cheng · Jun-Yan He · Kai Wang · Yuxiang Lin · Zheng Lian · Xiaojiang Peng · Alexander Hauptmann
HENASY: Learning to Assemble Scene-Entities for Interpretable Egocentric Video-Language Model
Khoa Vo · Thanh Phan · Kashu Yamazaki · Minh Tran · Ngan Le
Pre-trained Text-to-Image Diffusion Models Are Versatile Representation Learners for Control
Gunshi Gupta · Karmesh Yadav · Yarin Gal · Dhruv Batra · Zsolt Kira · Cong Lu · Tim G. J. Rudner
Meteor: Mamba-based Traversal of Rationale for Large Language and Vision Models
Byung-Kwan Lee · Chae Won Kim · Beomchan Park · Yong Man Ro
Visual Sketchpad: Sketching as a Visual Chain of Thought for Multimodal Language Models
Yushi Hu · Weijia Shi · Xingyu Fu · Dan Roth · Mari Ostendorf · Luke Zettlemoyer · Noah Smith · Ranjay Krishna
WATT: Weight Average Test Time Adaptation of CLIP
David OSOWIECHI · Mehrdad Noori · Gustavo Vargas Hakim · Moslem Yazdanpanah · Ali Bahri · Milad Cheraghalikhani · Sahar Dastani · Farzad Beizaei · Ismail Ayed · Christian Desrosiers
GenArtist: Multimodal LLM as an Agent for Unified Image Generation and Editing
Zhenyu Wang · Aoxue Li · Zhenguo Li · Xihui Liu
MoVA: Adapting Mixture of Vision Experts to Multimodal Context
ZHUOFAN ZONG · Bingqi Ma · Dazhong Shen · Guanglu Song · Hao Shao · DONGZHI JIANG · Hongsheng Li · Yu Liu
One-to-Normal: Anomaly Personalization for Few-shot Anomaly Detection
Yiyue Li · Shaoting Zhang · Kang Li · Qicheng Lao
Matryoshka Query Transformer for Large Vision-Language Models
Wenbo Hu · Zi-Yi Dou · Liunian Li · Amita Kamath · Nanyun Peng · Kai-Wei Chang
MaVen: An Effective Multi-granularity Hybrid Visual Encoding Framework for Multimodal Large Language Model
Chaoya Jiang · Hongrui Jia · Haiyang Xu · Wei Ye · Mengfan Dong · Ming Yan · Ji Zhang · Fei Huang · Shikun Zhang
LLMs Can Evolve Continually on Modality for \mathbb{X}-Modal Reasoning
Jiazuo Yu · Haomiao Xiong · Lu Zhang · Haiwen Diao · Yunzhi Zhuge · Lanqing Hong · Dong Wang · Huchuan Lu · You He · Long Chen
SongCreator: Lyrics-based Universal Song Generation
Shun Lei · Yixuan Zhou · Boshi Tang · Max W. Y. Lam · Feng Liu · Hangyu Liu · Jingcheng Wu · Shiyin Kang · Zhiyong Wu · Helen Meng
Bridge the Modality and Capability Gaps in Vision-Language Model Selection
Chao Yi · Yuhang He · De-Chuan Zhan · Han-Jia Ye
Make-it-Real: Unleashing Large Multimodal Model for Painting 3D Objects with Realistic Materials
Ye Fang · Zeyi Sun · Tong Wu · Jiaqi Wang · Ziwei Liu · Gordon Wetzstein · Dahua Lin
Grounding Multimodal Large Language Models in Actions
Andrew Szot · Bogdan Mazouze · Harsh Agrawal · R Devon Hjelm · Zsolt Kira · Alexander Toshev
DeiSAM: Segment Anything with Deictic Prompting
Hikaru Shindo · Manuel Brack · Gopika Sudhakaran · Devendra S Dhami · Patrick Schramowski · Kristian Kersting
Federated Learning from Vision-Language Foundation Models: Theoretical Analysis and Method
Bikang Pan · Wei Huang · Ye Shi
MambaTree: Tree Topology is All You Need in State Space Model
Yicheng Xiao · Lin Song · shaoli huang · Jiangshan Wang · Siyu Song · Yixiao Ge · Xiu Li · Ying Shan
OpenDign: Open-World Point Cloud Understanding with Depth-Aligned Images
Ye Mao · JUNPENG JING · Krystian Mikolajczyk
Tackling Uncertain Correspondences for Multi-Modal Entity Alignment
Liyi Chen · Ying Sun · Shengzhe Zhang · Yuyang Ye · Wei Wu · Hui Xiong
RestoreAgent: Autonomous Image Restoration Agent via Multimodal Large Language Models
Haoyu Chen · Wenbo Li · Jinjin Gu · Jingjing Ren · Sixiang Chen · Tian Ye · Renjing Pei · Kaiwen Zhou · Fenglong Song · Lei Zhu
Octopus: A Multi-modal LLM with Parallel Recognition and Sequential Understanding
Chuyang Zhao · YuXin Song · Junru Chen · KANG RONG · Haocheng Feng · Gang Zhang · Shufan Ji · Jingdong Wang · Errui Ding · Yifan Sun
An eye for an ear: zero-shot audio description leveraging an image captioner with audio-visual token distribution matching
Hugo Malard · Michel Olvera · Stéphane Lathuilière · Slim Essid
Voila-A: Aligning Vision-Language Models with User's Gaze Attention
Kun Yan · Zeyu Wang · Lei Ji · Yuntao Wang · Nan Duan · Shuai Ma
Cambrian-1: A Fully Open, Vision-Centric Exploration of Multimodal LLMs
Peter Tong · Ellis Brown · Penghao Wu · Sanghyun Woo · Adithya Jairam Vedagiri IYER · Sai Charitha Akula · Shusheng Yang · Jihan Yang · Manoj Middepogu · Ziteng Wang · Xichen Pan · Rob Fergus · Yann LeCun · Saining Xie
Does Video-Text Pretraining Help Open-Vocabulary Online Action Detection?

qingsong zhao · Yi Wang · Jilan Xu · Yinan He · Zifan Song · Limin Wang · Yu Qiao · Cairong Zhao
Exploring the Role of Large Language Models in Prompt Encoding for Diffusion Models
Bingqi Ma · ZHUOFAN ZONG · Guanglu Song · Hongsheng Li · Yu Liu
INQUIRE: A Natural World Text-to-Image Retrieval Benchmark
Edward Vendrow · Omiros Pantazis · Alexander Shepard · Gabriel Brostow · Kate Jones · Oisín Mac Aodha · Sara Beery · Grant Van Horn
DreamClear: High-Capacity Real-World Image Restoration with Privacy-Safe Dataset Curation
Yuang Ai · Xiaoqiang Zhou · Huaibo Huang · Xiaotian Han · Zhengyu Chen · Quanzeng You · Hongxia Yang
ControlMLLM: Training-Free Visual Prompt Learning for Multimodal Large Language Models
Mingrui Wu · Xinyue Cai · Jiayi Ji · Jiale Li · Oucheng Huang · Gen Luo · Hao Fei · GUANNAN JIANG · Xiaoshuai Sun · Rongrong Ji
Web-Scale Visual Entity Recognition: An LLM-Driven Data Approach
Mathilde Caron · Alireza Fathi · Cordelia Schmid · Ahmet Iscen
CharXiv: Charting Gaps in Realistic Chart Understanding in Multimodal LLMs
Zirui Wang · Mengzhou Xia · Luxi He · Howard Chen · Yitao Liu · Richard Zhu · Kaiqu Liang · Xindi Wu · Haotian Liu · Sadhika Malladi · Chevalier · Sanjeev Arora · Danqi Chen
OmniTokenizer: A Joint Image-Video Tokenizer for Visual Generation
Junke Wang · Yi Jiang · Zehuan Yuan · BINGYUE PENG · Zuxuan Wu · Yu-Gang Jiang
Enhancing Large Vision Language Models with Self-Training on Image Comprehension
Yihe Deng · Pan Lu · Fan Yin · Ziniu Hu · Sheng Shen · Quanquan Gu · James Zou · Kai-Wei Chang · Wei Wang
UMFC: Unsupervised Multi-Domain Feature Calibration for Vision-Language Models
Jiachen Liang · RuiBing Hou · Minyang Hu · Hong Chang · Shiguang Shan · Xilin Chen
SocialGPT: Prompting LLMs for Social Relation Reasoning via Greedy Segment Optimization
Wanhua Li · Zibin Meng · Jiawei Zhou · Donglai Wei · Chuang Gan · Hanspeter Pfister
Architect: Generating Vivid and Interactive 3D Scenes with Hierarchical 2D Inpainting
Yian Wang · Xiaowen Qiu · Jiageng Liu · Zhehuan Chen · Jiting Cai · Yufei Wang · Tsun-Hsuan Johnson Wang · Zhou Xian · Chuang Gan
Customized Multiple Clustering via Multi-Modal Subspace Proxy Learning
Jiawei Yao · Qi Qian · Juhua Hu
Multi-modal Situated Reasoning in 3D Scenes
Xiongzun Linghu · Xuesong Niu · Jiangyong Huang · Xiaojian (Shawn) Ma · Baoxiong Jia · Siyuan Huang
Kaleido Diffusion: Improving Conditional Diffusion Models with Autoregressive Latent Modeling
Jiatao Gu · Ying Shen · Shuangfei Zhai · Yizhe Zhang · Navdeep Jaitly · Joshua Susskind
Slot-VLM: Object-Event Slots for Video-Language Modeling
Jiaqi Xu · Cuiling Lan · Wenxuan Xie · Xuejin Chen · Yan Lu
Token Merging for Training-Free Semantic Binding in Text-to-Image Synthesis
Taihang Hu · Linxuan Li · Joost van de Weijer · Hongcheng Gao · Fahad Shahbaz Khan · Jian Yang · Ming-Ming Cheng · KAI WANG · Yaxing Wang
Needle In A Multimodal Haystack
Weiyun Wang · Shuibo Zhang · Yiming Ren · Yuchen Duan · Tiantong Li · Shuo Liu · Mengkang Hu · Zhe Chen · Kaipeng Zhang · Lewei Lu · Xizhou Zhu · Ping Luo · Yu Qiao · Jifeng Dai · Wenqi Shao · Wenhai Wang
UQE: A Query Engine for Unstructured Databases
Hanjun Dai · Bethany Wang · Xingchen Wan · Bo Dai · Sherry Yang · Azade Nova · Pengcheng Yin · Mangpo Phothilimthana · Charles Sutton · Dale Schuurmans
OneRef: Unified One-tower Expression Grounding and Segmentation with Mask Referring Modeling
Linhui Xiao · Xiaoshan Yang · Fang Peng · Yaowei Wang · Changsheng Xu
Uni-Med: A Unified Medical Generalist Foundation Model For Multi-Task Learning Via Connector-MoE
Xun Zhu · Ying Hu · Fanbin Mo · Miao Li · Ji Wu
Compositional 3D-aware Video Generation with LLM Director
Hanxin Zhu · Tianyu He · Anni Tang · Junliang Guo · Zhibo Chen · Jiang Bian
Vision Foundation Model Enables Generalizable Object Pose Estimation
Kai Chen · Yiyao Ma · Xingyu Lin · Stephen James · Jianshu Zhou · Yun-Hui Liu · Pieter Abbeel · DOU QI
KptLLM: Unveiling the Power of Large Language Model for Keypoint Comprehension
Jie Yang · Wang ZENG · Sheng Jin · Lumin Xu · Wentao Liu · Chen Qian · Ruimao Zhang
DeepStack: Deeply Stacking Visual Tokens is Surprisingly Simple and Effective for LMMs
Lingchen Meng · Jianwei Yang · Rui Tian · Xiyang Dai · Zuxuan Wu · Jianfeng Gao · Yu-Gang Jiang
DenseFusion-1M: Merging Vision Experts for Comprehensive Multimodal Perception
Xiaotong Li · Fan Zhang · Haiwen Diao · Yuezhe Wang · Xinlong Wang · LINGYU DUAN
Homology Consistency Constrained Efficient Tuning for Vision-Language Models
Huatian Zhang · Lei Zhang · Yongdong Zhang · Zhendong Mao
EAGLE: Efficient Adaptive Geometry-based Learning in Cross-view Understanding
Thanh-Dat Truong · Utsav Prabhu · Dongyi Wang · Bhiksha Raj · Susan Gauch · Jeyamkondan Subbiah · Khoa Luu

MetaUAS: Universal Anomaly Segmentation with One-Prompt Meta-Learning
Bin-Bin Gao
TOPA: Extending Large Language Models for Video Understanding via Text-Only Pre-Alignment
Wei Li · Hehe Fan · Yongkang Wong · Mohan Kankanhalli · Yi Yang
CALVIN: Improved Contextual Video Captioning via Instruction Tuning
Gowthami Somepalli · Arkabandhu Chowdhury · Jonas Geiping · Ronen Basri · Tom Goldstein · David Jacobs
CuMo: Scaling Multimodal LLM with Co-Upcycled Mixture-of-Experts
Jiachen Li · Xinyao Wang · Sijie Zhu · Chia-Wen Kuo · Lu XU · Fan Chen · Jitesh Jain · Humphrey Shi · Longyin Wen
Magnet: We Never Know How Text-to-Image Diffusion Models Work, Until We Learn How Vision-Language Models Function
Chenyi Zhuang · Ying Hu · Pan Gao
AP-Adapter: Improving Generalization of Automatic Prompts on Unseen Text-to-Image Diffusion Models
Yuchen Fu · Zhiwei Jiang · Yuliang Liu · Cong Wang · Zexuan Deng · Zhaoling Chen · Qing Gu
A Unified Debiasing Approach for Vision-Language Models across Modalities and Tasks
Hoin Jung · Taeuk Jang · Xiaoqian Wang
BoostAdapter: Improving Vision-Language Test-Time Adaptation via Regional Bootstrapping
Taolin Zhang · Jinpeng Wang · Hang Guo · Tao Dai · Bin Chen · Shu-Tao Xia
Text-Guided Attention is All You Need for Zero-Shot Robustness in Vision-Language Models
Lu Yu · Haiyang Zhang · Changsheng Xu
AdaNeg: Adaptive Negative Proxy Guided OOD Detection with Vision-Language Models
Yabin Zhang · Lei Zhang
Towards Flexible Visual Relationship Segmentation
Fangrui Zhu · Jianwei Yang · Huaizu Jiang
Reasoning and prompting
Self-playing Adversarial Language Game Enhances LLM Reasoning
Pengyu Cheng · Tianhao Hu · Han Xu · Zhisong Zhang · Yong Dai · Lei Han · nan du · Xiaolong Li
ReST-MCTS*: LLM Self-Training via Process Reward Guided Tree Search
Dan Zhang · Sining Zhoubian · Ziniu Hu · Yisong Yue · Yuxiao Dong · Jie Tang
On the Worst Prompt Performance of Large Language Models
Bowen Cao · Deng Cai · Zhisong Zhang · Yuexian Zou · Wai Lam
Symbolic Regression with a Learned Concept Library
Arya Grayeli · Atharva Sehgal · Omar Costilla Reyes · Miles Cranmer · Swarat Chaudhuri
AutoPSV: Automated Process-Supervised Verifier
Jianqiao Lu · Zhiyang Dou · Hongru WANG · Zeyu Cao · Jianbo Dai · Yunlong Feng · Zhijiang Guo
Toward Self-Improvement of LLMs via Imagination, Searching, and Criticizing
Ye Tian · Baolin Peng · Linfeng Song · Lifeng Jin · Dian Yu · Lei Han · Haitao Mi · Dong Yu
Advancing Tool-Augmented Large Language Models: Integrating Insights from Errors in Inference Trees
SIJIA CHEN · Yibo Wang · Yi-Feng Wu · Qingguo Chen · Zhao Xu · Weihua Luo · Kaifu Zhang · Lijun Zhang
From Unstructured Data to In-Context Learning: Exploring What Tasks Can Be Learned and When
Kevin Christian Wibisono · Yixin Wang
Where does In-context Learning \ Happen in Large Language Models?
Suzanna Sia · David Mueller · Kevin Duh
Proving Theorems Recursively
Haiming Wang · Huajian Xin · Zhengying Liu · Wenda Li · Yinya Huang · Jianqiao Lu · Zhicheng Yang · Jing Tang · Jian Yin · Zhenguo Li · Xiaodan Liang
Probing the Decision Boundaries of In-context Learning in Large Language Models
Siyan Zhao · Tung Nguyen · Aditya Grover
Information Re-Organization Improves Reasoning in Large Language Models
Xiaoxia Cheng · Zeqi Tan · Wei Xue · Weiming Lu
Do LLMs dream of elephants (when told not to)? Latent concept association and associative memory in transformers
Yibo Jiang · Goutham Rajendran · Pradeep Ravikumar · Bryon Aragam
Diffusion of Thought: Chain-of-Thought Reasoning in Diffusion Language Models
Jiacheng Ye · Shansan Gong · Liheng Chen · Lin Zheng · Jiahui Gao · Han Shi · Chuan Wu · Xin Jiang · Zhenguo Li · Wei Bi · Lingpeng Kong
Leveraging Hallucinations to Reduce Manual Prompt Dependency in Promptable Segmentation
Jian Hu · Jiayi Lin · Junchi Yan · Shaogang Gong
Buffer of Thoughts: Thought-Augmented Reasoning with Large Language Models
Ling Yang · Zhaochen Yu · Tianjun Zhang · Shiyi Cao · Minkai Xu · Wentao Zhang · Joseph Gonzalez · Bin CUI
SemCoder: Training Code Language Models with Comprehensive Semantics Reasoning
Yangruibo Ding · Jinjun Peng · Marcus Min · Gail Kaiser · Junfeng Yang · Baishakhi Ray

Instance-adaptive Zero-shot Chain-of-Thought Prompting
Xiaosong Yuan · Chen Shen · Shaotian Yan · Xiaofeng Zhang · Liang Xie · Wenxiao Wang · Renchu Guan · Ying Wang · Jieping Ye
Can Language Models Perform Robust Reasoning in Chain-of-thought Prompting with Noisy Rationales?
Zhanke Zhou · Rong Tao · Jianing Zhu · Yiwen Luo · Zengmao Wang · Bo Han
Doing Experiments and Revising Rules with Natural Language and Probabilistic Reasoning
Top Piriyaakulkij · Cassidy Langenfeld · Tuan Anh Le · Kevin Ellis
Iteration Head: A Mechanistic Study of Chain-of-Thought
Vivien Cabannes · Charles Arnal · Wassim Bouaziz · Xingyu Yang · Francois Charton · Julia Kempe
Enhancing Zero-Shot Vision Models by Label-Free Prompt Distribution Learning and Bias Correcting
Xingyu Zhu · Beier Zhu · Yi Tan · Shuo Wang · Yanbin Hao · Hanwang Zhang
LLM Processes: Numerical Predictive Distributions Conditioned on Natural Language
James Requeima · John Bronskill · Dami Choi · Richard Turner · David Duvenaud
Grokking of Implicit Reasoning in Transformers: A Mechanistic Journey to the Edge of Generalization
Boshi Wang · Xiang Yue · Yu Su · Huan Sun
Towards a Theoretical Understanding of the 'Reversal Curse' via Training Dynamics
Hanlin Zhu · Baihe Huang · Shaolun Zhang · Michael Jordan · Jiantao Jiao · Yuandong Tian · Stuart J Russell
Calibrating Reasoning in Language Models with Internal Consistency
Zhihui Xie · Jizhou Guo · Tong Yu · Shuai Li
In-Context Learning State Vector with Inner and Momentum Optimization
Dongfang Li · zhenyu liu · Xinshuo Hu · Zetian Sun · Baotian Hu · Min Zhang
Chain of Preference Optimization: Improving Chain-of-Thought Reasoning in LLMs
Xuan Zhang · Chao Du · Tianyu Pang · Qian Liu · Wei Gao · Min Lin
Adaptable Logical Control for Large Language Models
Honghua Zhang · Po-Nien Kung · Masahiro Yoshida · Guy Van den Broeck · Nanyun Peng
Learning to Reason via Program Generation, Emulation, and Search
Nathaniel Weir · Muhammad Khalifa · Linlu Qiu · Orion Weller · Peter Clark
Enhancing In-Context Learning Performance with just SVD-Based Weight Pruning: A Theoretical Perspective
Xinhao Yao · Xiaolin Hu · Shenzi Yang · Yong Liu
Uncertainty of Thoughts: Uncertainty-Aware Planning Enhances Information Seeking in LLMs
Zhiyuan Hu · Chumin Liu · Xidong Feng · Yilun Zhao · See-Kiong Ng · Anh Tuan Luu · Junxian He · Pang Wei Koh · Bryan Hooi
Unlocking the Capabilities of Thought: A Reasoning Boundary Framework to Quantify and Optimize Chain-of-Thought
Qiguang Chen · Libo Qin · Jiaqi Wang · Jingxuan Zhou · Wanxiang Che
Embedding Trajectory for Out-of-Distribution Detection in Mathematical Reasoning
Yiming Wang · Pei Zhang · Baosong Yang · Derek Wong · Zhuosheng Zhang · Rui Wang
Can Language Models Learn to Skip Steps?
Tengxiao Liu · Qipeng Guo · Xiangkun Hu · Cheng Jiayang · Yue Zhang · Xipeng Qiu · Zheng Zhang
TACT: Advancing Complex Aggregative Reasoning with Information Extraction Tools
Avi Caciularu · Alon Jacovi · Eyal Ben-David · Sasha Goldshtein · Tal Schuster · Jonathan Herzig · Gal Elidan · Amir Globerson
On the Inductive Bias of Stacking Towards Improving Reasoning
Nikunj Saunshi · Stefani Karp · Shankar Krishnan · Sobhan Miryoosefi · Sashank Jakkam Reddi · Sanjiv Kumar
Delving into the Reversal Curse: How Far Can Large Language Models Generalize?
Zhengkai Lin · Zhihang Fu · Kai Liu · Liang Xie · Binbin Lin · Wenxiao Wang · Deng Cai · Yue Wu · Jieping Ye
Cascade Speculative Drafting for Even Faster LLM Inference
Ziyi Chen · Xiaocong Yang · Jiacheng Lin · Chenkai Sun · Kevin Chang · Jie Huang
Causal language modeling can elicit search and reasoning capabilities on logic puzzles
Kulin Shah · Nishanth Dikkala · Xin Wang · Rina Panigrahy
Mind's Eye of LLMs: Visualization-of-Thought Elicits Spatial Reasoning in Large Language Models
Wenshan Wu · Shaoguang Mao · Yadong Zhang · Yan Xia · Li Dong · Lei Cui · Furu Wei
BERTs are Generative In-Context Learners
David Samuel
On the Noise Robustness of In-Context Learning for Text Generation
hongfu gao · Feipeng Zhang · Wenyu Jiang · Jun Shu · Feng Zheng · Hongxin Wei
Data Mixture Inference Attack: BPE Tokenizers Reveal Training Data Compositions
Jonathan Hayase · Alisa Liu · Yejin Choi · Sewoong Oh · Noah Smith
Limits of Transformer Language Models on Learning to Compose Algorithms
Jonathan Thomm · Giacomo Camposampiero · Aleksandar Terzic · Michael Hersche · Bernhard Schölkopf · Abbas Rahimi
GTBench: Uncovering the Strategic Reasoning Capabilities of LLMs via Game-Theoretic Evaluations
Jinhao Duan · Renming Zhang · James Diffenderfer · Bhavya Kailkhura · Lichao Sun · Elias Stengel-Eskin · Mohit Bansal · Tianlong Chen · Kaidi Xu
On the Power of Decision Trees in Auto-Regressive Language Modeling
Yulu Gan · Tomer Galanti · Tomaso Poggio · Eran Malach

Global Lyapunov functions: a long-standing open problem in mathematics, with symbolic transformers
Alberto Alfarano · Francois Charton · Amaury Hayat
Discovery of the Hidden World with Large Language Models
Chenxi Liu · Yongqiang Chen · Tongliang Liu · Mingming Gong · James Cheng · Bo Han · Kun Zhang
In-Context Learning with Representations: Contextual Generalization of Trained Transformers
Tong Yang · Yu Huang · Yingbin Liang · Yuejie Chi
Mixture of Demonstrations for In-Context Learning
Song Wang · Zihan Chen · Chengshuai Shi · Cong Shen · Jundong Li
Language Generation in the Limit
Jon Kleinberg · Sendhil Mullainathan
Chain-of-Thought Unfaithfulness as Disguised Accuracy
Ana Marasovic · Nathan Stringham · Oliver Bentham
Unveiling Causal Reasoning in Large Language Models: Reality or Mirage?
Haoang Chi · He Li · Wenjing Yang · Feng Liu · Long Lan · Xiaoguang Ren · Tongliang Liu · Bo Han
SELF-DISCOVER: Large Language Models Self-Compose Reasoning Structures
Pei Zhou · Jay Pujara · Xiang Ren · Xinyun Chen · Heng-Tze Cheng · Quoc V Le · Ed Chi · Denny Zhou · Swaroop Mishra · Huaixiu (Steven) Zheng
Plan-on-Graph: Self-Correcting Adaptive Planning of Large Language Model on Knowledge Graphs
Liyi Chen · Panrong Tong · Zhongming Jin · Ying Sun · Jieping Ye · Hui Xiong
Provably Transformers Harness Multi-Concept Word Semantics for Efficient In-Context Learning
Dake Bu · Wei Huang · Andi Han · Atsushi Nitanda · Taiji Suzuki · Qingfu Zhang · Hau-San Wong
Concentrate Attention: Towards Domain-Generalizable Prompt Optimization for Language Models
Chengzhengxu Li · Xiaoming Liu · Zhaohan Zhang · Yichen Wang · Chen Liu · Yu Lan · Chao Shen
Meaningful Learning: Enhancing Abstract Reasoning in Large Language Models via Generic Fact Guidance
Kai Xiong · Xiao Ding · Ting Liu · Bing Qin · Dongliang Xu · Qing Yang · Hongtao Liu · Yixin Cao
MindMerger: Efficiently Boosting LLM Reasoning in non-English Languages
Zixian Huang · Wenhao Zhu · Gong Cheng · Lei Li · Fei Yuan
Optimized Feature Generation for Tabular Data via LLMs with Decision Tree Reasoning
Jaehyun Nam · Kyuyoung Kim · Seunghyuk Oh · Jihoon Tack · Jaehyung Kim · Jinwoo Shin
Self-Guiding Exploration for Combinatorial Problems
Zangir Iklassov · Yali Du · Farkhad Akimov · Martin Takac
Teach Better or Show Smarter? On Instructions and Exemplars in Automatic Prompt Optimization
Xingchen Wan · Ruoxi Sun · Hootan Nakhost · Sercan Arik
Prompt Optimization with EASE? Efficient Ordering-aware Automated Selection of Exemplars
Zhaoxuan Wu · Xiaoqiang Lin · Zhongxiang Dai · Wenyang Hu · Yao Shu · See-Kiong Ng · Patrick Jaillet · Bryan Kian Hsiang Low
Localized Zeroth-Order Prompt Optimization
Wenyang Hu · Yao Shu · Zongmin Yu · Zhaoxuan Wu · Xiaoqiang Lin · Zhongxiang Dai · See-Kiong Ng · Bryan Kian Hsiang Low
SpatialPIN: Enhancing Spatial Reasoning Capabilities of Vision-Language Models through Prompting and Interacting 3D Priors
Chenyang Ma · Kai Lu · Ta-Ying Cheng · Niki Trigoni · Andrew Markham
JiuZhang3.0: Efficiently Improving Mathematical Reasoning by Training Small Data Synthesis Models
Kun Zhou · Beichen Zhang · Jiapeng Wang · Zhipeng Chen · Xin Zhao · Jing Sha · Zhichao Sheng · Shijin Wang · Ji-Rong Wen
Mars: Situated Inductive Reasoning in an Open-World Environment
Xiaojuan Tang · Jiaqi Li · Yitao Liang · Song-Chun Zhu · Muhan Zhang · Zilong Zheng
The Evolution of Statistical Induction Heads: In-Context Learning Markov Chains
Ezra Edelman · Nikolaos Tsilivis · Benjamin Edelman · Eran Malach · Surbhi Goel
Enhancing Reasoning Capabilities of LLMs via Principled Synthetic Logic Corpus
Terufumi Morishita · Gaku Morio · Atsuki Yamaguchi · Yasuhiro Sogawa
AlphaMath Almost Zero: Process Supervision without Process
Guoxin Chen · Minpeng Liao · Chengxi Li · Kai Fan
ALPINE: Unveiling The Planning Capability of Autoregressive Learning in Language Models
Siwei Wang · Yifei Shen · Shi Feng · Haoran Sun · Shang-Hua Teng · Wei Chen
Many-Shot In-Context Learning
Rishabh Agarwal · Avi Singh · Lei Zhang · Bernd Bohnet · Luis Rosias · Stephanie Chan · Biao Zhang · Ankesh Anand · Zaheer Abbas · Azade Nova · John Co-Reyes · Eric Chu · Feryal Behbahani · Aleksandra Faust · Hugo Larochelle
Chain-of-Thought Reasoning Without Prompting
Xuezhi Wang · Denny Zhou
Connecting the Dots: LLMs can Infer and Verbalize Latent Structure from Disparate Training Data
Johannes Treutlein · Dami Choi · Jan Betley · Samuel Marks · Cem Anil · Roger Grosse · Owain Evans
Rule Extrapolation in Language Modeling: A Study of Compositional Generalization on OOD Prompts
Anna Mészáros · Szilvia Ujváry · Wieland Brendel · Patrik Reizinger · Ferenc Huszar

Does Reasoning Emerge? Examining the Probabilities of Causation in Large Language Models
Javier Gonzalez · Aditya Nori
Co-occurrence is not Factual Association in Language Models
Xiao Zhang · Miao Li · Ji Wu
MAmmoTH2: Scaling Instructions from the Web
Xiang Yue · Tianyu Zheng · Ge Zhang · Wenhui Chen
Microstructures and Accuracy of Graph Recall by Large Language Models
Yanbang Wang · Hejie Cui · Jon Kleinberg
Chain of Thoughtlessness? An Analysis of CoT in Planning
Kaya Stechly · Karthik Valmeekam · Subbarao Kambhampati
Order-Independence Without Fine Tuning
Reid McIlroy-Young · Katrina Brown · Conlan Olson · Linjun Zhang · Cynthia Dwork
Decompose, Analyze and Rethink: Solving Intricate Problems with Human-like Reasoning Cycle
Shangzi Xue · Zhenya Huang · Jiayu Liu · Xin Lin · Yuting Ning · Binbin Jin · Xin Li · Qi Liu
Can large language models explore in-context?
Akshay Krishnamurthy · Keegan Harris · Dylan J Foster · Cyril Zhang · Aleksandrs Slivkins
LLMs as Zero-shot Graph Learners: Alignment of GNN Representations with LLM Token Embeddings
Duo Wang · Yuan Zuo · Fengzhi Li · Junjie Wu
Divide-and-Conquer Meets Consensus: Unleashing the Power of Functions in Code Generation
Jingchang Chen · Hongxuan Tang · Zheng Chu · Qianglong Chen · Zekun Wang · Ming Liu · Bing Qin
Surveys
Algorithmic progress in language models
Wing Hin Anson Ho · Tamay Besiroglu · Ege Erdil · Zifan Guo · David Owen · Robi Rahman · David Atkinson · Neil Thompson · Jaime Sevilla
Questioning the Survey Responses of Large Language Models
Ricardo Dominguez-Olmedo · Moritz Hardt · Celestine Mendler-Dünnér
Text Detection
Rethinking LLM Memorization through the Lens of Adversarial Compression
Avi Schwarzschild · Zhili Feng · Pratyush Maini · Zachary Lipton · J. Zico Kolter
Membership Inference Attacks against Large Vision-Language Models
Zhan Li · Yongtao Wu · Yihang Chen · Francesco Tonin · Elias Abad Rocamora · Volkan Cevher
DLAD: Improving Logits-based Detector without Logits from Black-box LLMs
Cong Zeng · Shengkun Tang · Xianjun Yang · Yuanzhou Chen · Yiyou Sun · Zhiqiang Xu · Yao Li · Haifeng Chen · Wei Cheng · Dongkuan (DK) Xu
Inevitable Trade-off between Watermark Strength and Speculative Sampling Efficiency for Language Models
Zhengmian Hu · Heng Huang
WaterMax: breaking the LLM watermark detectability-robustness-quality trade-off
Eva Giboulot · Teddy Furon
LLM Dataset Inference: Did you train on my dataset?
Pratyush Maini · Hengrui Jia · Nicolas Papernot · Adam Dziedzić
Edit Distance Robust Watermarks via Indexing Pseudorandom Codes
Noah Golowich · Ankur Moitra
BiScope: AI-generated Text Detection by Checking Memorization of Preceding Tokens
Hanxi Guo · Siyuan Cheng · Xiaolong Jin · Zhuo Zhang · Kaiyuan Zhang · Guan hong Tao · Guangyu Shen · Xiangyu Zhang
Watermarking Makes Language Models Radioactive
Tom Sander · Pierre Fernandez · Alain Durmus · Matthijs Douze · Teddy Furon
HuRef: HUMAN-REadable Fingerprint for Large Language Models
Boyi Zeng · Lizheng Wang · Yuncong Hu · Yi Xu · Chenghu Zhou · Xinbing Wang · Yu Yu · Zhouhan Lin
DeTeCtive: Detecting AI-generated Text via Multi-Level Contrastive Learning
Xun Guo · Yongxin He · Shan Zhang · Ting Zhang · Wanquan Feng · Haibin Huang · Chongyang Ma
Segmenting Watermarked Texts From Language Models
Xingchi Li · Guanxun Li · Xianyang Zhang
Bileve: Securing Text Provenance in Large Language Models Against Spoofing with Bi-level Signature
Tong Zhou · Xuandong Zhao · Xiaolin Xu · Shaolei Ren
DPIC: Decoupling Prompt and Intrinsic Characteristics for LLM Generated Text Detection
XIAO YU · Yang Qi · Kejiang Chen · Guoqiang Chen · Xi Yang · Pengyuan Zhu · Xiuwei Shang · Weiming Zhang · Nenghai Yu
Others
AdaNeg: Adaptive Negative Proxy Guided OOD Detection with Vision-Language Models

Yabin Zhang · Lei Zhang
Towards Flexible Visual Relationship Segmentation
Fangrui Zhu · Jianwei Yang · Huaizu Jiang
AdaNeg: Adaptive Negative Proxy Guided OOD Detection with Vision-Language Models
Yabin Zhang · Lei Zhang
Towards Flexible Visual Relationship Segmentation
Fangrui Zhu · Jianwei Yang · Huaizu Jiang
Biologically-Inspired Learning Model for Instructed Vision
Roy Abel · Shimon Ullman
Super Consistency of Neural Network Landscapes and Learning Rate Transfer
Lorenzo Noci · Alexandru Meterez · Thomas Hofmann · Antonio Orvieto
Are More LLM Calls All You Need? Towards the Scaling Properties of Compound AI Systems
Lingjiao Chen · Jared Quincy Davis · Boris Hanin · Peter Bailis · Ion Stoica · Matei A Zaharia · James Zou
Linking In-context Learning in Transformers to Human Episodic Memory
Ji-An Li · Corey Zhou · Marcus Benna · Marcelo G Mattar
Ensemble Learning for Heterogeneous Large Language Models with Deep Parallel Collaboration
Yichong Huang · Xiaocheng Feng · Baohang Li · Yang Xiang · Hui Wang · Ting Liu · Bing Qin
xLSTM: Extended Long Short-Term Memory
Maximilian Beck · Korbinian Pöppel · Markus Spanring · Andreas Auer · Oleksandra Prudnikova · Michael Kopp · Günter Klambauer · Johannes Brandstetter · Sepp Hochreiter
The Expressive Capacity of State Space Models: A Formal Language Perspective
Yash Sarrof · Yana Veitsman · Michael Hahn
Stealth edits to large language models
Oliver Sutton · Qinghua Zhou · Wei Wang · Desmond Higham · Alexander N Gorban · Alexander Bastounis · Ivan Tyukin
Data-Driven Discovery of Dynamical Systems in Pharmacology using Large Language Models
Samuel Holt · Zhaozhi Qian · Tennison Liu · Jim Weatherall · Mihaela van der Schaar
Should We Really Edit Language Models? On the Evaluation of Edited Language Models
Qi Li · Xiang Liu · Zhenheng Tang · Peijie Dong · Zeyu Li · Xinglin Pan · Xiaowen Chu
SSDM: Scalable Speech Dysfluency Modeling
Jiachen Lian · Xuanru Zhou · Zoe Ezzes · Jet Vonk · Brittany Morin · David Paul Baquirin · Zachary Miller · Maria Luisa Gorno Tempini · Gopala Anumanchipalli
Evidence of Learned Look-Ahead in a Chess-Playing Neural Network
Erik Jenner · Shreyas Kapur · Vasil Georgiev · Cameron Allen · Scott Emmons · Stuart J Russell
Evaluating the World Model Implicit in a Generative Model
Keyon Vafa · Justin Chen · Ashesh Rambachan · Jon Kleinberg · Sendhil Mullainathan
No Free Delivery Service: Epistemic limits of passive data collection in complex social systems
Maximilian Nickel
Universal Sample Coding
Szymon Kobus · Tze-Yang Tung · Deniz Gunduz
Length Optimization in Conformal Prediction
Shayan Kiyani · George J. Pappas · Hamed Hassani
Zipfian Whitening
Sho Yokoi · Han Bao · Hiroto Kurita · Hidetoshi Shimodaira
MomentumSMoE: Integrating Momentum into Sparse Mixture of Experts
Rachel S.Y. Teo · Tan Nguyen
The Closeness of In-Context Learning and Weight Shifting for Softmax Regression
Shuai Li · Zhao Song · Yu Xia · Tong Yu · Tianyi Zhou
Membership Inference Attacks against Fine-tuned Large Language Models via Self-prompt Calibration
Wenjie Fu · Huandong Wang · Chen Gao · Guanghua Liu · Yong Li · Tao Jiang
Theoretical Analysis of Weak-to-Strong Generalization
Hunter Lang · David Sontag · Aravindan Vijayaraghavan
Fisher Flow Matching for Generative Modeling over Discrete Data
Oscar Davis · Samuel Kessler · Mircea Petrache · Ismail Ceylan · Michael Bronstein · Joey Bose
PANORAMIA: Privacy Auditing of Machine Learning Models without Retraining
Mishaal Kazmi · Hadrien Lautreite · Alireza Akbari · Qiaoyue Tang · Mauricio Soroco · Tao Wang · Sébastien Gambs · Mathias Lé cuyer
On the Scalability of GNNs for Molecular Graphs
Maciej Sypetkowski · Frederik Wenkel · Farimah Poursafaei · Nia Dickson · Karush Suri · Philip Fradkin · Dominique Beaini
Universal Neural Functionals
Allan Zhou · Chelsea Finn · James Harrison
A Full-duplex Speech Dialogue Scheme Based On Large Language Model

<i>Peng Wang · Songshuo Lu · Yaohua Tang · Sijie Yan · Wei Xia · Yuanjun Xiong</i>
<i>Fair Wasserstein Coresets</i>
<i>Zikai Xiong · Niccolo Dalmaso · Shubham Sharma · Freddy Lecue · Daniele Magazzeni · Vamsi Potluru · Tucker Balch · Manuela Veloso</i>
<i>Bayesian-guided Label Mapping for Visual Reprogramming</i>
<i>Chengyi Cai · Zesheng Ye · Lei Feng · Jianzhong Qi · Feng Liu</i>
<i>Are Language Models Actually Useful for Time Series Forecasting?</i>
<i>Mingtian Tan · Mike Merrill · Vinayak Gupta · Tim Althoff · Tom Hartvigsen</i>
<i>AutoSurvey: Large Language Models Can Automatically Write Surveys</i>
<i>Yidong Wang · Qi Guo · Wenjin Yao · Hongbo Zhang · Xin Zhang · Zhen Wu · Meishan Zhang · Xinyu Dai · Min zhang · Qingsong Wen · Wei Ye · Shikun Zhang · Yue Zhang</i>
<i>Understanding Scaling Laws with Statistical and Approximation Theory for Transformer Neural Networks on Intrinsically Low-dimensional Data</i>
<i>Alexander Havrilla · Wenjing Liao</i>
<i>FlowLLM: Flow Matching for Material Generation with Large Language Models as Base Distributions</i>
<i>Anuroop Sriram · Benjamin K Miller · Ricky T. Q. Chen · Brandon Wood</i>
<i>An Analysis of Tokenization: Transformers under Markov Data</i>
<i>Nived Rajaraman · Jiantao Jiao · Kannan Ramchandran</i>