arXiv cs.AI Latest Papers (227 total)

Date: 2025-07-24

Towards Autonomous Sustainability Assessment via Multimodal AI Agents

Authors: Zhihan Zhang, Alexander Metzger, Yuxuan Mei, Felix H\"ahnlein, Zachary Englhardt, Tingyu Cheng, Gregory D. Abowd, Shwetak Patel, Adriana Schulz, Vikram Iyer

Abstract: arXiv:2507.17012v1 Announce Type: new Abstract: Interest in sustainability information has surged in recent years. However, the data required for a life cycle assessment (LCA) that maps the materials and processes from product manufacturing to disposal into environmental impacts (EI) are often unavailable. Here we reimagine conventional LCA by introducing multimodal AI agents that emulate interactions between LCA experts and stakeholders like product managers and engineers to calculate the cradle-to-gate (production) carbon emissions of electronic devices. The AI agents iteratively generate a detailed life-cycle inventory leveraging a custom data abstraction and software tools that extract information from online text and images from repair communities and government certifications. This approach reduces weeks or months of expert time to under one minute and closes data availability gaps while yielding carbon footprint estimates within 19% of expert LCAs with zero proprietary data. Additionally, we develop a method to directly estimate EI by comparing an input to a cluster of products with similar descriptions and known carbon footprints. This runs in 3 ms on a laptop with a MAPE of 12.28% on electronic products. Further, we develop a data-driven method to generate emission factors. We use the properties of an unknown material to represent it as a weighted sum of emission factors for similar materials. Compared to human experts picking the closest LCA database entry, this improves MAPE by 120.26%. We analyze the data and compute scaling of this approach and discuss its implications for future LCA workflows.

View Paper

New Mechanisms in Flex Distribution for Bounded Suboptimal Multi-Agent Path Finding

Authors: Shao-Hung Chan, Thomy Phan, Jiaoyang Li, Sven Koenig

Abstract: arXiv:2507.17054v1 Announce Type: new Abstract: Multi-Agent Path Finding (MAPF) is the problem of finding a set of collision-free paths, one for each agent in a shared environment. Its objective is to minimize the sum of path costs (SOC), where the path cost of each agent is defined as the travel time from its start location to its target location. Explicit Estimation Conflict-Based Search (EECBS) is the leading algorithm for bounded-suboptimal MAPF, with the SOC of the solution being at most a user-specified factor w away from optimal. EECBS maintains sets of paths and a lower bound LB on the optimal SOC. Then, it iteratively selects a set of paths whose SOC is at most $w \cdot LB$ and introduces constraints to resolve collisions. For each path in a set, EECBS maintains a lower bound on its optimal path that satisfies constraints. By finding an individually bounded-suboptimal path with cost at most a threshold of w times its lower bound, EECBS guarantees to find a bounded-suboptimal solution. To speed up EECBS, previous work uses flex distribution to increase the threshold. Though EECBS with flex distribution guarantees to find a bounded-suboptimal solution, increasing the thresholds may push the SOC beyond $w \cdot LB$, forcing EECBS to switch among different sets of paths instead of resolving collisions on a particular set of paths, and thus reducing efficiency. To address this issue, we propose Conflict-Based Flex Distribution that distributes flex in proportion to the number of collisions. We also estimate the delays needed to satisfy constraints and propose Delay-Based Flex Distribution. On top of that, we propose Mixed-Strategy Flex Distribution, combining both in a hierarchical framework. We prove that EECBS with our new flex distribution mechanisms is complete and bounded-suboptimal. Our experiments show that our approaches outperform the original (greedy) flex distribution.

View Paper

LoRA is All You Need for Safety Alignment of Reasoning LLMs

Authors: Yihao Xue, Baharan Mirzasoleiman

Abstract: arXiv:2507.17075v1 Announce Type: new Abstract: Reasoning LLMs have demonstrated remarkable breakthroughs in solving complex problems that were previously out of reach. To ensure LLMs do not assist with harmful requests, safety alignment fine-tuning is necessary in the post-training phase. However, safety alignment fine-tuning has recently been shown to significantly degrade reasoning abilities, a phenomenon known as the "Safety Tax". In this work, we show that using LoRA for SFT on refusal datasets effectively aligns the model for safety without harming its reasoning capabilities. This is because restricting the safety weight updates to a low-rank space minimizes the interference with the reasoning weights. Our extensive experiments across four benchmarks covering math, science, and coding show that this approach

produces highly safe LLMs -- with safety levels comparable to full-model fine-tuning -- without compromising their reasoning abilities. Additionally, we observe that LoRA induces weight updates with smaller overlap with the initial weights compared to full-model fine-tuning. We also explore methods that further reduce such overlap -- via regularization or during weight merging -- and observe some improvement on certain tasks. We hope this result motivates designing approaches that yield more consistent improvements in the reasoning-safety trade-off.

View Paper

HySafe-AI: Hybrid Safety Architectural Analysis Framework for AI Systems: A Case Study

Authors: Mandar Pitale, Jelena Frtunikj, Abhinaw Priyadershi, Vasu Singh, Maria Spence

Abstract: arXiv:2507.17118v1 Announce Type: new Abstract: AI has become integral to safety-critical areas like autonomous driving systems (ADS) and robotics. The architecture of recent autonomous systems are trending toward end-to-end (E2E) monolithic architectures such as large language models (LLMs) and vision language models (VLMs). In this paper, we review different architectural solutions and then evaluate the efficacy of common safety analyses such as failure modes and effect analysis (FMEA) and fault tree analysis (FTA). We show how these techniques can be improved for the intricate nature of the foundational models, particularly in how they form and utilize latent representations. We introduce HySAFE-AI, Hybrid Safety Architectural Analysis Framework for AI Systems, a hybrid framework that adapts traditional methods to evaluate the safety of AI systems. Lastly, we offer hints of future work and suggestions to guide the evolution of future AI safety standards.

View Paper

Improving LLMs' Generalized Reasoning Abilities by Graph Problems

Authors: Qifan Zhang, Nuo Chen, Zehua Li, Miao Peng, Jing Tang, Jia Li

Abstract: arXiv:2507.17168v1 Announce Type: new Abstract: Large Language Models (LLMs) have made remarkable strides in reasoning tasks, yet their performance often falters on novel and complex problems. Domain-specific continued pretraining (CPT) methods, such as those tailored for mathematical reasoning, have shown promise but lack transferability to broader reasoning tasks. In this work, we pioneer the use of Graph Problem Reasoning (GPR) to enhance the general reasoning capabilities of LLMs. GPR tasks, spanning pathfinding, network analysis, numerical computation, and topological reasoning, require sophisticated logical and relational reasoning, making them ideal for teaching diverse reasoning patterns. To achieve this, we introduce GraphPile, the first large-scale corpus specifically designed for CPT using GPR data. Spanning 10.9 billion tokens across 23 graph tasks, the dataset includes chain-of-thought, program-of-thought, trace of execution, and real-world graph data. Using GraphPile, we train GraphMind on popular base models Llama 3 and 3.1, as well as Gemma 2, achieving up to 4.9 percent higher accuracy in mathematical reasoning and up to 21.2 percent improvement in non-mathematical reasoning tasks such as logical and commonsense reasoning. By being the first to harness GPR for enhancing reasoning patterns and introducing the first dataset of its kind, our work bridges the gap between domain-specific pretraining and universal reasoning capabilities, advancing the adaptability and robustness of LLMs.

View Paper

Our Cars Can Talk: How IoT Brings AI to Vehicles

Authors: Amod Kant Agrawal

Abstract: arXiv:2507.17214v1 Announce Type: new Abstract: Bringing AI to vehicles and enabling them as sensing platforms is key to transforming maintenance from reactive to proactive. Now is the time to integrate AI copilots that speak both languages: machine and driver. This article offers a conceptual and technical perspective intended to spark interdisciplinary dialogue and guide future research and development in intelligent vehicle systems, predictive maintenance, and AI-powered user interaction.

View Paper

Agent Identity Evals: Measuring Agentic Identity

Authors: Elija Perrier, Michael Timothy Bennett

Abstract: arXiv:2507.17257v1 Announce Type: new Abstract: Central to agentic capability and trustworthiness of language model agents (LMAs) is the extent they maintain stable, reliable, identity over time. However, LMAs inherit pathologies from large language models (LLMs) (statelessness, stochasticity, sensitivity to prompts and linguistically-intermediation) which can undermine their identifiability, continuity, persistence and consistency. This attrition of identity can erode their reliability, trustworthiness and utility by interfering with their agentic capabilities such as reasoning, planning and action. To address these challenges, we introduce \textit{agent identity evals} (AIE), a rigorous, statistically-driven, empirical framework for measuring the degree to which an LMA system exhibit and maintain their agentic identity over time, including their capabilities, properties and ability to recover from state perturbations. AIE comprises a set of novel metrics which can integrate with other measures of performance, capability and agentic robustness to assist in the design of optimal LMA infrastructure and scaffolding such as memory and tools. We set out formal definitions and methods that can be applied at each stage of the LMA life-cycle, and worked examples of how to apply them.

View Paper

Students' Feedback Requests and Interactions with the SCRIPT Chatbot: Do They Get What They Ask For?

Authors: Andreas Scholl, Natalie Kiesler

Abstract: arXiv:2507.17258v1 Announce Type: new Abstract: Building on prior research on Generative AI (GenAI) and related tools for programming education, we developed SCRIPT, a chatbot based on ChatGPT-4o-mini, to support novice learners. SCRIPT allows for open-ended interactions and structured guidance through predefined prompts. We evaluated the tool via an experiment with 136 students from an introductory programming course at a large German university and analyzed how students interacted with SCRIPT while solving programming tasks with a focus on their feedback preferences. The results reveal that students' feedback requests seem to follow a specific sequence. Moreover, the chatbot responses aligned well with students' requested feedback types (in 75%), and it adhered to the system prompt constraints. These insights inform the design of GenAI-based learning support systems and highlight challenges in balancing guidance and flexibility in AI-assisted tools.

View Paper

Compliance Brain Assistant: Conversational Agentic AI for Assisting Compliance Tasks in Enterprise Environments

Authors: Shitong Zhu, Chenhao Fang, Derek Larson, Neel Reddy Pochareddy, Rajeev Rao, Sophie Zeng, Yanqing Peng, Wendy Summer, Alex Goncalves, Arya Pudota, Herve Robert

Abstract: arXiv:2507.17289v1 Announce Type: new Abstract: This paper presents Compliance Brain Assistant (CBA), a conversational, agentic AI assistant designed to boost the efficiency of daily compliance tasks for personnel in enterprise environments. To strike a good balance between response quality and latency, we design a user query router that can intelligently choose between (i) FastTrack mode: to handle simple requests that only need additional relevant context retrieved from knowledge corpora; and (ii) FullAgentic mode: to handle complicated requests that need composite actions and tool invocations to proactively discover context across various compliance artifacts, and/or involving other APIs/models for accommodating requests. A typical example would be to start with a user query, use its description to find a specific entity and then use the entity's information to query other APIs for curating and enriching the final AI response. Our experimental evaluations compared CBA against an out-of-the-box LLM on various real-world privacy/compliance-related queries targeting various personas. We found that CBA substantially improved upon the vanilla LLM's performance on metrics such as average keyword match rate (83.7% vs. 41.7%) and LLM-judge pass rate (82.0% vs. 20.0%). We also compared metrics for the full routing-based design against the `fast-track only` and `full-agentic` modes and found that it had a better average match-rate and pass-rate while keeping the run-time approximately the same. This finding validated our hypothesis that the routing mechanism leads to a good trade-off between the two worlds.

View Paper

Ctx2TrajGen: Traffic Context-Aware Microscale Vehicle Trajectories using Generative Adversarial Imitation Learning

Authors: Joobin Jin, Seokjun Hong, Gyeongseon Baek, Yeeun Kim, Byeongjoon Noh

Abstract: arXiv:2507.17418v1 Announce Type: new Abstract: Precise modeling of microscopic vehicle trajectories is critical for traffic behavior analysis and autonomous driving systems. We propose Ctx2TrajGen, a context-aware trajectory generation framework that synthesizes realistic urban driving behaviors using GAIL. Leveraging PPO and WGAN-GP, our model addresses nonlinear interdependencies and training instability inherent in microscopic settings. By explicitly conditioning on surrounding vehicles and road geometry, Ctx2TrajGen generates interaction-aware trajectories aligned with real-world context. Experiments on the drone-captured DRIFT dataset demonstrate superior performance over existing methods in terms of realism, behavioral diversity, and contextual fidelity, offering a robust solution to data scarcity and domain shift without simulation.

View Paper

An Uncertainty-Driven Adaptive Self-Alignment Framework for Large Language Models

Authors: Haoran Sun, Zekun Zhang, Shaoning Zeng

Abstract: arXiv:2507.17477v1 Announce Type: new Abstract: Large Language Models (LLMs) have demonstrated remarkable progress in instruction following and general-purpose reasoning. However, achieving high-quality alignment with human intent and safety norms without human annotations remains a fundamental challenge. In this work, we propose an Uncertainty-Driven Adaptive Self-Alignment (UDASA) framework designed to improve LLM alignment in a fully automated manner. UDASA first generates multiple responses for each input and quantifies output uncertainty across three dimensions: semantics, factuality, and value alignment. Based on these uncertainty scores, the framework constructs preference pairs and categorizes training samples into three stages, conservative, moderate, and exploratory, according to their uncertainty difference. The model is then optimized progressively across these stages. In addition, we conduct a series of preliminary studies to validate the core design assumptions and provide strong empirical motivation for the proposed framework. Experimental results show that UDASA outperforms existing alignment methods across multiple tasks, including harmlessness, helpfulness, truthfulness, and controlled sentiment generation, significantly improving model performance.

View Paper

LTLZinc: a Benchmarking Framework for Continual Learning and Neuro-Symbolic Temporal Reasoning

Authors: Luca Salvatore Lorello, Nikolaos Manginas, Marco Lippi, Stefano Melacci

Abstract: arXiv:2507.17482v1 Announce Type: new Abstract: Neuro-symbolic artificial intelligence aims to combine neural architectures with symbolic approaches that can represent knowledge in a human-interpretable formalism. Continual learning concerns with agents that expand their knowledge over time, improving their skills while avoiding to forget previously learned concepts. Most of the existing approaches for neuro-symbolic artificial intelligence are applied to static scenarios only, and the challenging setting where reasoning along the temporal dimension is necessary has been seldom explored. In this work we introduce LTLZinc, a benchmarking framework that can be used to generate datasets covering a variety of different problems, against which neuro-symbolic

and continual learning methods can be evaluated along the temporal and constraint-driven dimensions. Our framework generates expressive temporal reasoning and continual learning tasks from a linear temporal logic specification over MiniZinc constraints, and arbitrary image classification datasets. Fine-grained annotations allow multiple neural and neuro-symbolic training settings on the same generated datasets. Experiments on six neuro-symbolic sequence classification and four class-continual learning tasks generated by LTLZinc, demonstrate the challenging nature of temporal learning and reasoning, and highlight limitations of current state-of-the-art methods. We release the LTLZinc generator and ten ready-to-use tasks to the neuro-symbolic and continual learning communities, in the hope of fostering research towards unified temporal learning and reasoning frameworks.

View Paper

CQE under Epistemic Dependencies: Algorithms and Experiments (extended version)

Authors: Lorenzo Marconi, Flavia Ricci, Riccardo Rosati

Abstract: arXiv:2507.17487v1 Announce Type: new Abstract: We investigate Controlled Query Evaluation (CQE) over ontologies, where information disclosure is regulated by epistemic dependencies (EDs), a family of logical rules recently proposed for the CQE framework. In particular, we combine EDs with the notion of optimal GA censors, i.e. maximal sets of ground atoms that are entailed by the ontology and can be safely revealed. We focus on answering Boolean unions of conjunctive queries (BUCQs) with respect to the intersection of all optimal GA censors - an approach that has been shown in other contexts to ensure strong security guarantees with favorable computational behavior. First, we characterize the security of this intersection-based approach and identify a class of EDs (namely, full EDs) for which it remains safe. Then, for a subclass of EDs and for DL-Lite_R ontologies, we show that answering BUCQs in the above CQE semantics is in AC^0 in data complexity by presenting a suitable, detailed first-order rewriting algorithm. Finally, we report on experiments conducted in two different evaluation scenarios, showing the practical feasibility of our rewriting function.

View Paper

Automated Hybrid Grounding Using Structural and Data-Driven Heuristics

Authors: Alexander Beiser, Markus Hecher, Stefan Woltran

Abstract: arXiv:2507.17493v1 Announce Type: new Abstract: The grounding bottleneck poses one of the key challenges that hinders the widespread adoption of Answer Set Programming in industry. Hybrid Grounding is a step in alleviating the bottleneck by combining the strength of standard bottom-up grounding with recently proposed techniques where rule bodies are decoupled during grounding. However, it has remained unclear when hybrid grounding shall use body-decoupled grounding and when to use standard bottom-up grounding. In this paper, we address this issue by developing automated hybrid grounding: we introduce a splitting algorithm based on data-structural heuristics that detects when to use body-decoupled grounding and when standard grounding is beneficial. We base our heuristics on the structure of rules and an estimation procedure that incorporates the data of the instance. The experiments conducted on our prototypical implementation demonstrate promising results, which show an improvement on hard-to-ground scenarios, whereas on hard-to-solve instances we approach state-of-the-art performance.

View Paper

Can One Domain Help Others? A Data-Centric Study on Multi-Domain Reasoning via Reinforcement Learning

Authors: Yu Li, Zhuoshi Pan, Honglin Lin, Mengyuan Sun, Conghui He, Lijun Wu

Abstract: arXiv:2507.17512v1 Announce Type: new Abstract: Reinforcement Learning with Verifiable Rewards (RLVR) has emerged as a powerful paradigm for enhancing the reasoning capabilities of LLMs. Existing research has predominantly concentrated on isolated reasoning domains such as mathematical problem-solving, coding tasks, or logical reasoning. However, real world reasoning scenarios inherently demand an integrated application of multiple cognitive skills. Despite this, the interplay among these reasoning skills under reinforcement learning remains poorly understood. To bridge this gap, we present a systematic investigation of multi-domain reasoning within the RLVR framework, explicitly focusing on three primary domains: mathematical reasoning, code generation, and logical puzzle solving. We conduct a comprehensive study comprising four key components: (1) Leveraging the GRPO algorithm and the Qwen-2.5-7B model family, our study thoroughly evaluates the models' in-domain improvements and cross-domain generalization capabilities when trained on single-domain datasets. (2) Additionally, we examine the intricate interactions including mutual enhancements and conflicts that emerge during combined cross-domain training. (3) To further understand the influence of SFT on RL, we also analyze and compare performance differences between base and instruct models under identical RL configurations. (4) Furthermore, we

delve into critical RL training details, systematically exploring the impacts of curriculum learning strategies, variations in reward design, and language-specific factors. Through extensive experiments, our results offer significant insights into the dynamics governing domain interactions, revealing key factors influencing both specialized and generalizable reasoning performance. These findings provide valuable guidance for optimizing RL methodologies to foster comprehensive, multi-domain reasoning capabilities in LLMs.

View Paper

TAI Scan Tool: A RAG-Based Tool With Minimalistic Input for Trustworthy AI Self-Assessment

Authors: Athanasios Davvetas, Xenia Ziouvelou, Ypatia Dami, Alexis Kaponis, Konstantina Giouvanopoulou, Michael Papademas

Abstract: arXiv:2507.17514v1 Announce Type: new Abstract: This paper introduces the TAI Scan Tool, a RAG-based TAI self-assessment tool with minimalistic input. The current version of the tool supports the legal TAI assessment, with a particular emphasis on facilitating compliance with the AI Act. It involves a two-step approach with a pre-screening and an assessment phase. The assessment output of the system includes insight regarding the risk-level of the AI system according to the AI Act, while at the same time retrieving relevant articles to aid with compliance and notify on their obligations. Our qualitative evaluation using use-case scenarios yields promising results, correctly predicting risk levels while retrieving relevant articles across three distinct semantic groups. Furthermore, interpretation of results shows that the tool's reasoning relies on comparison with the setting of high-risk systems, a behaviour attributed to their deployment requiring careful consideration, and therefore frequently presented within the AI Act.

View Paper

Constructing Ophthalmic MLLM for Positioning-diagnosis Collaboration Through Clinical Cognitive Chain Reasoning

Authors: Xinyao Liu, Diping Song

Abstract: arXiv:2507.17539v1 Announce Type: new Abstract: Multimodal large language models (MLLMs) demonstrate significant potential in the field of medical diagnosis. However, they face critical challenges in specialized domains such as ophthalmology, particularly the fragmentation of annotation granularity and inconsistencies in clinical reasoning logic, which hinder precise cross-modal understanding. This paper introduces FundusExpert, an ophthalmology-specific MLLM with integrated positioning-diagnosis reasoning capabilities, along with FundusGen, a dataset constructed through the intelligent Fundus-Engine system. Fundus-Engine automates localization and leverages MLLM-based semantic expansion to integrate global disease classification, local object detection, and fine-grained feature analysis within a single fundus image. Additionally, by constructing a clinically aligned cognitive chain, it guides the model to generate interpretable reasoning paths. FundusExpert, fine-tuned with instruction data from FundusGen, achieves the best performance in ophthalmic question-answering tasks, surpassing the average accuracy of the 40B MedRegA by 26.6%. It also excels in zero-shot report generation tasks, achieving a clinical consistency of 77.0%, significantly outperforming GPT-4o's 47.6%. Furthermore, we reveal a scaling law between data quality and model capability ($L \propto N^{0.068}$), demonstrating that the cognitive alignment annotations in FundusGen enhance data utilization efficiency. By integrating region-level localization with diagnostic reasoning chains, our work develops a scalable, clinically-aligned MLLM and explores a pathway toward bridging the visual-language gap in specific MLLMs. Our project can be found at https://github.com/MeteorElf/FundusExpert.

View Paper

Simulating multiple human perspectives in socio-ecological systems using large language models

Authors: Yongchao Zeng, Calum Brown, Ioannis Kyriakou, Ronja Hotz, Mark Rounsevell

Abstract: arXiv:2507.17680v1 Announce Type: new Abstract: Understanding socio-ecological systems requires insights from diverse stakeholder perspectives, which are often hard to access. To enable alternative, simulation-based exploration of different stakeholder perspectives, we develop the HoPeS (Human-Oriented Perspective Shifting) modelling framework. HoPeS employs agents powered by large language models (LLMs) to represent various stakeholders; users can step into the agent roles to experience perspectival differences. A simulation protocol serves as a "scaffold" to streamline multiple perspective-taking simulations, supporting users in reflecting on, transitioning between, and integrating across perspectives. A prototype system is developed to demonstrate HoPeS in the context of institutional dynamics and land use change, enabling both narrative-driven and numerical experiments. In an illustrative experiment, a user successively adopts the perspectives of a system observer and a researcher - a role that analyses data from the embedded land use model to inform evidence-based decision-making for other LLM agents representing various institutions. Despite the user's effort to recommend technically sound policies, discrepancies persist between the policy recommendation and implementation due to stakeholders' competing advocacies, mirroring real-world misalignment between researcher and policymaker perspectives. The user's reflection highlights the subjective feelings of frustration and disappointment as a researcher, especially due to the challenge of maintaining political neutrality while attempting to gain political influence. Despite this, the user exhibits high motivation to experiment with alternative narrative framing strategies, suggesting the system's potential in exploring different perspectives. Further system and protocol refinement are likely to enable new forms of interdisciplinary collaboration in socio-ecological simulations.

View Paper

Symbiotic Agents: A Novel Paradigm for Trustworthy AGI-driven Networks

Authors: Ilias Chatzistefanidis, Navid Nikaein

Abstract: arXiv:2507.17695v1 Announce Type: new Abstract: Large Language Model (LLM)-based autonomous agents are expected to play a vital role in the evolution of 6G networks, by empowering real-time decision-making related to management and service provisioning to end-users. This shift facilitates the transition from a specialized intelligence approach, where artificial intelligence (AI) algorithms handle isolated tasks, to artificial general intelligence (AGI)-driven networks, where agents possess broader reasoning capabilities and can manage diverse network functions. In this paper, we introduce a novel agentic paradigm that combines LLMs with real-time optimization algorithms towards Trustworthy AI, defined as symbiotic agents. Optimizers at the LLM's input-level provide bounded uncertainty steering for numerically precise tasks, whereas output-level optimizers supervised by the LLM enable adaptive real-time control. We design and implement two novel agent types including: (i) Radio Access Network optimizers, and (ii) multi-agent negotiators for Service-Level Agreements (SLAs). We further propose an end-to-end architecture for AGI networks and evaluate it on a 5G testbed capturing channel fluctuations from moving vehicles. Results show that symbiotic agents reduce decision errors fivefold compared to standalone LLM-based agents, while smaller language models (SLM) achieve similar accuracy with a 99.9% reduction in GPU resource overhead and in near-real-time loops of 82 ms. A multi-agent demonstration for collaborative RAN on the real-world testbed highlights significant flexibility in service-level agreement and resource allocation, reducing RAN over-utilization by approximately 44%. Drawing on our findings and open-source implementations, we introduce the symbiotic paradigm as the foundation for next-generation, AGI-driven networks-systems designed to remain adaptable, efficient, and trustworthy even as LLMs advance.

View Paper

Thinking Isn't an Illusion: Overcoming the Limitations of Reasoning Models via Tool Augmentations

Authors: Zhao Song, Song Yue, Jiahao Zhang

Abstract: arXiv:2507.17699v1 Announce Type: new Abstract: Large Reasoning Models (LRMs) have become a central focus in today's large language model (LLM) research, where models are designed to output a step-by-step thinking process before arriving at a final answer to handle complex reasoning tasks. Despite their promise, recent empirical studies (e.g., [Shojaee et al., 2025] from Apple) suggest that this thinking process may not actually enhance reasoning ability, where LLMs without explicit reasoning actually outperform LRMs on tasks with low or high complexity. In this work, we revisit these findings and investigate whether the limitations of LRMs persist when tool augmentations are introduced. We incorporate two types of tools, Python interpreters and scratchpads, and evaluate three representative LLMs and their LRM counterparts on Apple's benchmark reasoning puzzles. Our results show that, with proper tool use, LRMs consistently outperform their non-reasoning counterparts across all levels of task complexity. These findings challenge the recent narrative that reasoning is an illusion and highlight the potential of tool-augmented LRMs for solving complex problems.

View Paper

Online Submission and Evaluation System Design for Competition Operations

Authors: Zhe Chen, Daniel Harabor, Ryan Hechnenberger, Nathan R. Sturtevant

Abstract: arXiv:2507.17730v1 Announce Type: new Abstract: Research communities have developed benchmark datasets across domains to compare the performance of algorithms and techniques However, tracking the progress in these research areas is not easy, as publications appear in different venues at the same time, and many of them claim to represent the state-of-the-art. To address this, research communities often organise periodic competitions to evaluate the performance of various algorithms and techniques, thereby tracking advancements in the field. However, these competitions pose a significant operational burden. The organisers must manage and evaluate a large volume of submissions. Furthermore, participants typically develop their solutions in diverse environments, leading to compatibility issues during the evaluation of their submissions. This paper presents an online competition system that automates the submission and evaluation process for a competition. The competition system allows organisers to manage large numbers of submissions efficiently, utilising isolated environments to evaluate submissions. This system has already been used successfully for several competitions, including the Grid-Based Pathfinding Competition and the League of Robot Runners competition.

View Paper

Bridging Robustness and Generalization Against Word Substitution Attacks in NLP via the Growth Bound Matrix Approach

Authors: Mohammed Bouri, Adnane Saoud

Abstract: arXiv:2507.10330v1 Announce Type: cross Abstract: Despite advancements in Natural Language Processing (NLP), models remain vulnerable to adversarial attacks, such as synonym substitutions. While prior work has focused on improving robustness for feed-forward and convolutional architectures, the robustness of recurrent networks and modern state space models (SSMs), such as S4, remains understudied. These architectures pose unique challenges due to their sequential processing and complex parameter dynamics. In this paper, we introduce a novel regularization technique based on Growth Bound Matrices (GBM) to improve NLP model robustness by reducing the impact of input perturbations on model outputs. We focus on computing the GBM for three architectures: Long Short-Term Memory (LSTM), State Space models (S4), and Convolutional Neural Networks (CNN). Our method aims to (1) enhance resilience against word substitution attacks, (2) improve generalization on clean text, and (3) providing the first systematic analysis of SSM (S4) robustness. Extensive experiments across multiple architectures and benchmark datasets demonstrate that our method improves adversarial robustness by up to 8.8% over existing baselines. These results highlight the effectiveness of our approach, outperforming several state-of-the-art methods in adversarial defense. Codes are available at https://github.com/BouriMohammed/GBM

View Paper

Explainable Vulnerability Detection in C/C++ Using Edge-Aware Graph Attention Networks

Authors: Radowanul Haque, Aftab Ali, Sally McClean, Naveed Khan

Abstract: arXiv:2507.16540v1 Announce Type: cross Abstract: Detecting security vulnerabilities in source code remains challenging, particularly due to class imbalance in real-world datasets where vulnerable functions are under-represented. Existing learning-based methods often optimise for recall, leading to high false positive rates and reduced usability in development workflows. Furthermore, many approaches lack explainability, limiting their integration into security workflows. This paper presents ExplainVulD, a graph-based framework for vulnerability detection in C/C++ code. The method constructs Code Property Graphs and represents nodes using dual-channel embeddings that capture both semantic and structural information. These are processed by an edge-aware attention mechanism that incorporates edge-type embeddings to distinguish among program relations. To address class imbalance, the model is trained using class-weighted cross-entropy loss. ExplainVulD achieves a mean accuracy of 88.25 percent and an F1 score of 48.23 percent across 30 independent runs on the ReVeal dataset. These results represent relative improvements of 4.6 percent in accuracy and 16.9 percent in F1 score compared to the ReVeal model, a prior learning-based method. The framework also outperforms static analysis tools, with relative gains of 14.0 to 14.1 percent in accuracy and 132.2 to 201.2 percent in F1 score. Beyond improved detection performance, ExplainVulD produces explainable outputs by identifying the most influential code regions within each function, supporting transparency and trust in security triage.

View Paper

Disaster Informatics after the COVID-19 Pandemic: Bibliometric and Topic Analysis based on Large-scale Academic Literature

Authors: Ngan Tran, Haihua Chen, Ana Cleveland, Yuhan Zhou

Abstract: arXiv:2507.16820v1 Announce Type: cross Abstract: This study presents a comprehensive bibliometric and topic analysis of the disaster informatics literature published between January 2020 to September 2022. Leveraging a large-scale corpus and advanced techniques such as pre-trained language models and generative AI, we identify the most active countries, institutions, authors, collaboration networks, emergent topics, patterns among the most significant topics, and shifts in research priorities spurred by the COVID-19 pandemic. Our findings highlight (1) countries that were most impacted by the COVID-19 pandemic were also among the most active, with each country having specific research interests, (2) countries and institutions within the same region or share a common language tend to collaborate, (3) top active authors tend to form close partnerships with one or two key partners, (4) authors typically specialized in one or two specific topics, while institutions had more diverse interests across

several topics, and (5) the COVID-19 pandemic has influenced research priorities in disaster informatics, placing greater emphasis on public health. We further demonstrate that the field is converging on multidimensional resilience strategies and cross-sectoral data-sharing collaborations or projects, reflecting a heightened awareness of global vulnerability and interdependency. Collecting and quality assurance strategies, data analytic practices, LLM-based topic extraction and summarization approaches, and result visualization tools can be applied to comparable datasets or solve similar analytic problems. By mapping out the trends in disaster informatics, our analysis offers strategic insights for policymakers, practitioners, and scholars aiming to enhance disaster informatics capacities in an increasingly uncertain and complex risk landscape.

View Paper

A Query-Aware Multi-Path Knowledge Graph Fusion Approach for Enhancing Retrieval-Augmented Generation in Large Language Models

Authors: Qikai Wei, Huansheng Ning, Chunlong Han, Jianguo Ding

Abstract: arXiv:2507.16826v1 Announce Type: cross Abstract: Retrieval Augmented Generation (RAG) has gradually emerged as a promising paradigm for enhancing the accuracy and factual consistency of content generated by large language models (LLMs). However, existing RAG studies primarily focus on retrieving isolated segments using similarity-based matching methods, while overlooking the intrinsic connections between them. This limitation hampers performance in RAG tasks. To address this, we propose QMKGF, a Query-Aware Multi-Path Knowledge Graph Fusion Approach for Enhancing Retrieval Augmented Generation. First, we design prompt templates and employ general-purpose LLMs to extract entities and relations, thereby generating a knowledge graph (KG) efficiently. Based on the constructed KG, we introduce a multi-path subgraph construction strategy that incorporates one-hop relations, multi-hop relations, and importance-based relations, aiming to improve the semantic relevance between the retrieved documents and the user query. Subsequently, we designed a query-aware attention reward model that scores subgraph triples based on their semantic relevance to the query. Then, we select the highest score subgraph and enrich subgraph with additional triples from other subgraphs that are highly semantically relevant to the query. Finally, the entities, relations, and triples within the updated subgraph are utilised to expand the original query, thereby enhancing its semantic representation and improving the quality of LLMs' generation. We evaluate QMKGF on the SQuAD, IIRC, Culture, HotpotQA, and MuSiQue datasets. On the HotpotQA dataset, our method achieves a ROUGE-1 score of 64.98\%, surpassing the BGE-Rerank approach by 9.72 percentage points (from 55.26\% to 64.98\%). Experimental results demonstrate the effectiveness and superiority of the QMKGF approach.

View Paper

You Don't Bring Me Flowers: Mitigating Unwanted Recommendations Through Conformal Risk Control

Authors: Giovanni De Toni, Erasmo Purificato, Emilia G\'omez, Bruno Lepri, Andrea Passerini, Cristian Consonni

Abstract: arXiv:2507.16829v1 Announce Type: cross Abstract: Recommenders are significantly shaping online information consumption. While effective at personalizing content, these systems increasingly face criticism for propagating irrelevant, unwanted, and even harmful recommendations. Such content degrades user satisfaction and contributes to significant societal issues, including misinformation, radicalization, and erosion of user trust. Although platforms offer mechanisms to mitigate exposure to undesired content, these mechanisms are often insufficiently effective and slow to adapt to users' feedback. This paper introduces an intuitive, model-agnostic, and distribution-free method that uses conformal risk control to provably bound unwanted content in personalized recommendations by leveraging simple binary feedback on items. We also address a limitation of traditional conformal risk control approaches, i.e., the fact that the recommender can provide a smaller set of recommended items, by leveraging implicit feedback on consumed items to expand the recommendation set while ensuring robust risk mitigation. Our experimental evaluation on data coming from a popular online video-sharing platform demonstrates that our approach ensures an effective and controllable reduction of unwanted recommendations with minimal effort. The source code is available here: https://github.com/geektoni/mitigating-harm-recsys.

View Paper

Towards Robust Speech Recognition for Jamaican Patois Music Transcription

Authors: Jordan Madden, Matthew Stone, Dimitri Johnson, Daniel Geddez

Abstract: arXiv:2507.16834v1 Announce Type: cross Abstract: Although Jamaican Patois is a widely spoken language, current speech recognition systems perform poorly on Patois music, producing inaccurate captions that limit accessibility and hinder downstream applications. In this work, we take a data-centric approach to this problem by curating more than 40 hours of manually transcribed Patois music. We use this dataset to fine-tune state-of-the-art automatic speech recognition (ASR) models, and use the results to develop scaling laws for the performance of Whisper models on Jamaican Patois audio. We hope that this work will have a positive impact on the accessibility of Jamaican Patois music and the future of Jamaican Patois language modeling.

View Paper

Segmentation-free Goodness of Pronunciation

Authors: Xinwei Cao, Zijian Fan, Torbj{\o}rn Svendsen, Giampiero Salvi

Abstract: arXiv:2507.16838v1 Announce Type: cross Abstract: Mispronunciation detection and diagnosis (MDD) is a significant part in modern computer aided language learning (CALL) systems. Within MDD, phoneme-level pronunciation assessment is key to helping L2 learners improve their pronunciation. However, most systems are based on a form of goodness of pronunciation (GOP) which requires pre-segmentation of speech into phonetic units. This limits the accuracy of these methods and the possibility to use modern CTC-based acoustic models for their evaluation. In this study, we first propose self-alignment GOP (GOP-SA) that enables the use of CTC-trained ASR models for MDD. Next, we define a more general alignment-free method that takes all possible alignments of the target phoneme into account (GOP-AF). We give a theoretical account of our definition of GOP-AF, an implementation that solves potential numerical issues as well as a proper normalization which makes the method applicable with acoustic models with different peakiness over time. We provide extensive experimental results on the CMU Kids and Speechocean762 datasets comparing the different definitions of our methods, estimating the dependency of GOP-AF on the peakiness of the acoustic models and on the amount of context around the target phoneme. Finally, we compare our methods with recent studies over the Speechocean762 data showing that the feature vectors derived from the proposed method achieve state-of-the-art results on phoneme-level pronunciation assessment.

View Paper

CASPER: Contrastive Approach for Smart Ponzi Scheme Detecter with More Negative Samples

Authors: Weijia Yang, Tian Lan, Leyuan Liu, Wei Chen, Tianqing Zhu, Sheng Wen, Xiaosong Zhang

Abstract: arXiv:2507.16840v1 Announce Type: cross Abstract: The rapid evolution of digital currency trading, fueled by the integration of blockchain technology, has led to both innovation and the emergence of smart Ponzi schemes. A smart Ponzi scheme is a fraudulent investment operation in smart contract that uses funds from new investors to pay returns to earlier investors. Traditional Ponzi scheme detection methods based on deep learning typically rely on fully supervised models, which require large amounts of labeled data. However, such data is often scarce, hindering effective model training. To address this challenge, we propose a novel contrastive learning framework, CASPER (Contrastive Approach for Smart Ponzi detectER with more negative samples), designed to enhance smart Ponzi scheme detection in blockchain transactions. By leveraging contrastive learning techniques, CASPER can learn more effective representations of smart contract source code using unlabeled datasets, significantly reducing both operational costs and system complexity. We evaluate CASPER on the XBlock dataset, where it outperforms the baseline by 2.3% in F1 score when trained with 100% labeled data. More impressively, with only 25% labeled data, CASPER achieves an F1 score nearly 20% higher than the baseline under identical experimental conditions. These results highlight CASPER's potential for effective and cost-efficient detection of smart Ponzi schemes, paving the way for scalable fraud detection solutions in the future.

View Paper

Weak Supervision Techniques towards Enhanced ASR Models in Industry-level CRM Systems

Authors: Zhongsheng Wang, Sijie Wang, Jia Wang, Yung-I Liang, Yuxi Zhang, Jiamou Liu

Abstract: arXiv:2507.16843v1 Announce Type: cross Abstract: In the design of customer relationship management (CRM) systems, accurately identifying customer types and offering personalized services are key to enhancing customer satisfaction and loyalty. However, this process faces the challenge of discerning customer voices and intentions, and general pre-trained automatic speech recognition (ASR) models make it difficult to effectively address industry-specific speech recognition tasks. To address this issue, we innovatively proposed a solution for fine-tuning industry-specific ASR models, which significantly improved the performance of the fine-tuned ASR models in industry applications. Experimental results show that our method substantially improves the crucial auxiliary role of the ASR model in industry CRM systems, and this approach has also been adopted in actual industrial applications.

View Paper

Dynamic Simulation Framework for Disinformation Dissemination and Correction With Social Bots

Authors: Boyu Qiao, Kun Li, Wei Zhou, Songlin Hu

Abstract: arXiv:2507.16848v1 Announce Type: cross Abstract: In the human-bot symbiotic information ecosystem, social bots play key roles in spreading and correcting disinformation. Understanding their influence is essential for risk control and better governance. However, current studies often rely on simplistic user and network modeling, overlook the dynamic behavior of bots, and lack quantitative evaluation of correction strategies. To fill these gaps, we propose MADD, a Multi Agent based framework for Disinformation Dissemination. MADD constructs a more realistic propagation network by integrating the Barabasi Albert Model for scale free topology and the Stochastic Block Model for community structures, while designing node attributes based on real world user data. Furthermore, MADD incorporates both malicious and legitimate bots, with their controlled dynamic participation allows for quantitative analysis of correction strategies. We evaluate MADD using individual and group level metrics. We experimentally verify the real world consistency of MADD user attributes and network structure, and we simulate the dissemination of six disinformation topics, demonstrating the differential effects of fact based and narrative based correction strategies.

[View Paper]

Post-Disaster Affected Area Segmentation with a Vision Transformer (ViT)-based EVAP Model using Sentinel-2 and Formosat-5 Imagery

Authors: Yi-Shan Chu, Hsuan-Cheng Wei

Abstract: arXiv:2507.16849v1 Announce Type: cross Abstract: We propose a vision transformer (ViT)-based deep learning framework to refine disaster-affected area segmentation from remote sensing imagery, aiming to support and enhance the Emergent Value Added Product (EVAP) developed by the Taiwan Space Agency (TASA). The process starts with a small set of manually annotated regions. We then apply principal component analysis (PCA)-based feature space analysis and construct a confidence index (CI) to expand these labels, producing a weakly supervised training set. These expanded labels are then used to train ViT-based encoder-decoder models with multi-band inputs from Sentinel-2 and Formosat-5 imagery. Our architecture supports multiple decoder variants and multi-stage loss strategies to improve performance under limited supervision. During the evaluation, model predictions are compared with higher-resolution EVAP output to assess spatial coherence and segmentation consistency. Case studies on the 2022 Poyang Lake drought and the 2023 Rhodes wildfire demonstrate that our framework improves the smoothness and reliability of segmentation results, offering a scalable approach for disaster mapping when accurate ground truth is unavailable.

View Paper

Toward a Real-Time Framework for Accurate Monocular 3D Human Pose Estimation with Geometric Priors

Authors: Mohamed Adjel (LAAS)

Abstract: arXiv:2507.16850v1 Announce Type: cross Abstract: Monocular 3D human pose estimation remains a challenging and ill-posed problem, particularly in real-time settings and unconstrained environments. While direct imageto-3D approaches require large annotated datasets and heavy models, 2D-to-3D lifting offers a more lightweight and flexible alternative-especially when enhanced with prior knowledge. In this work, we propose a framework that combines real-time 2D keypoint detection with geometry-aware 2D-to-3D lifting, explicitly leveraging known camera intrinsics and subject-specific anatomical priors. Our approach builds on recent advances in self-calibration and biomechanically-constrained inverse kinematics to generate large-scale, plausible 2D-3D training pairs from MoCap and synthetic datasets. We discuss how these ingredients can enable fast, personalized, and accurate 3D pose estimation from monocular images without requiring specialized hardware. This proposal aims to foster discussion on bridging data-driven learning and model-based priors to improve accuracy, interpretability, and deployability of 3D human motion capture on edge devices in the wild.

View Paper

SynthCTI: LLM-Driven Synthetic CTI Generation to enhance MITRE Technique Mapping

Authors: \'Alvaro Ruiz-R\'odenas, Jaime Pujante S\'aez, Daniel Garc\'ia-Algora, Mario Rodr\'iguez B\'ejar, Jorge Blasco, Jos\'e Luis Hern\'andez-Ramos

Abstract: arXiv:2507.16852v1 Announce Type: cross Abstract: Cyber Threat Intelligence (CTI) mining involves extracting structured insights from unstructured threat data, enabling organizations to understand and respond to evolving adversarial behavior. A key task in CTI mining is mapping threat descriptions to MITRE ATT\&CK techniques. However, this process is often performed manually, requiring expert knowledge and substantial effort. Automated approaches face two major challenges: the scarcity of high-quality labeled CTI data and class imbalance, where many techniques have very few examples. While domain-specific Large Language Models (LLMs) such as SecureBERT have shown improved performance, most recent work focuses on model architecture rather than addressing the data limitations. In this work, we present SynthCTI, a data augmentation framework designed to generate high-quality synthetic CTI sentences for underrepresented MITRE ATT\&CK techniques. Our method uses a clustering-based strategy to extract semantic context from training data and guide an LLM in producing synthetic CTI sentences that are lexically diverse and semantically faithful. We evaluate SynthCTI on two publicly available CTI datasets, CTI-to-MITRE and TRAM, using LLMs with different capacity. Incorporating synthetic data leads to consistent macro-F1 improvements: for example, ALBERT improves from 0.35 to 0.52 (a relative gain of 48.6\%), and SecureBERT reaches 0.6558 (up from 0.4412). Notably, smaller models augmented with SynthCTI outperform larger models trained without augmentation, demonstrating the value of data generation methods for building efficient and effective CTI classification systems.

View Paper

CLAMP: Contrastive Learning with Adaptive Multi-loss and Progressive Fusion for Multimodal Aspect-Based Sentiment Analysis

Authors: Xiaoqiang He

Abstract: arXiv:2507.16854v1 Announce Type: cross Abstract: Multimodal aspect-based sentiment analysis(MABSA) seeks to identify aspect terms within paired image-text data and determine their fine grained sentiment polarities, representing a fundamental task for improving the effectiveness of applications such as product review systems and public opinion monitoring. Existing methods face challenges such as cross modal alignment noise and insufficient consistency in fine-grained representations. While global modality alignment methods often overlook the connection between aspect terms and their corresponding local visual regions, bridging the representation gap between text and images remains a challenge. To address these limitations, this paper introduces an end to end Contrastive Learning framework with Adaptive Multi-loss and Progressive Attention Fusion(CLAMP). The framework is composed of three novel modules: Progressive Attention Fusion network, Multi-task Contrastive Learning, and

Adaptive Multi-loss Aggregation. The Progressive Attention Fusion network enhances fine-grained alignment between textual features and image regions via hierarchical, multi-stage cross modal interactions, effectively suppressing irrelevant visual noise. Secondly, multi-task contrastive learning combines global modal contrast and local granularity alignment to enhance cross modal representation consistency. Adaptive Multi-loss Aggregation employs a dynamic uncertainty based weighting mechanism to calibrate loss contributions according to each task's uncertainty, thereby mitigating gradient interference. Evaluation on standard public benchmarks demonstrates that CLAMP consistently outperforms the vast majority of existing state of the art methods.

View Paper

SIA: Enhancing Safety via Intent Awareness for Vision-Language Models

Authors: Youngjin Na, Sangheon Jeong, Youngwan Lee

Abstract: arXiv:2507.16856v1 Announce Type: cross Abstract: As vision-language models (VLMs) are increasingly deployed in real-world applications, new safety risks arise from the subtle interplay between images and text. In particular, seemingly innocuous inputs can combine to reveal harmful intent, leading to unsafe model responses. Despite increasing attention to multimodal safety, previous approaches based on post hoc filtering or static refusal prompts struggle to detect such latent risks, especially when harmfulness emerges only from the combination of inputs. We propose SIA (Safety via Intent Awareness), a training-free prompt engineering framework that proactively detects and mitigates harmful intent in multimodal inputs. SIA employs a three-stage reasoning process: (1) visual abstraction via captioning, (2) intent inference through few-shot chain-of-thought prompting, and (3) intent-conditioned response refinement. Rather than relying on predefined rules or classifiers, SIA dynamically adapts to the implicit intent inferred from the image-text pair. Through extensive experiments on safety-critical benchmarks including SIUO, MM-SafetyBench, and HoliSafe, we demonstrate that SIA achieves substantial safety improvements, outperforming prior methods. Although SIA shows a minor reduction in general reasoning accuracy on MMStar, the corresponding safety gains highlight the value of intent-aware reasoning in aligning VLMs with human-centric values.

View Paper

Leveraging multi-source and heterogeneous signals for fatigue detection

Authors: Luobin Cui, Yanlai Wu, Tang Ying, Weikai Li

Abstract: arXiv:2507.16859v1 Announce Type: cross Abstract: Fatigue detection plays a critical role in safety-critical applications such as aviation, mining, and long-haul transport. However, most existing methods rely on high-end sensors and controlled environments, limiting their applicability in real world settings. This paper formally defines a practical yet underexplored problem setting for real world fatigue detection, where systems operating with context-appropriate sensors aim to leverage knowledge from differently instrumented sources including those using impractical sensors deployed in controlled environments. To tackle this challenge, we propose a heterogeneous and multi-source fatigue detection framework that adaptively utilizes the available modalities in the target domain while benefiting from the diverse configurations present in source domains. Our experiments, conducted using a realistic field-deployed sensor setup and two publicly available datasets, demonstrate the practicality, robustness, and improved generalization of our approach, paving the practical way for effective fatigue monitoring in sensor-constrained scenarios.

View Paper

Look Before You Fuse: 2D-Guided Cross-Modal Alignment for Robust 3D Detection

Authors: Xiang Li

Abstract: arXiv:2507.16861v1 Announce Type: cross Abstract: Integrating LiDAR and camera inputs into a unified Bird's-Eye-View (BEV) representation is crucial for enhancing 3D perception capabilities of autonomous vehicles. However, current methods are often affected by misalignment between camera and LiDAR features. This misalignment leads to inaccurate depth supervision in camera branch and erroneous fusion during cross-modal feature aggregation. The root cause of this misalignment lies in projection errors, stemming from minor extrinsic calibration inaccuracies and rolling shutter effect of LiDAR during vehicle motion. In this work, our key insight is that these projection errors are predominantly concentrated at object-background boundaries, which are readily identified by 2D detectors. Based on this, our main motivation is to utilize 2D object priors to pre-align cross-modal features before fusion. To address local misalignment, we propose Prior Guided Depth Calibration (PGDC), which leverages 2D priors to correct local misalignment and preserve correct cross-modal feature pairs. To resolve global misalignment, we introduce Discontinuity

Aware Geometric Fusion (DAGF) to process calibrated results from PGDC, suppressing noise and explicitly enhancing sharp transitions at object-background boundaries. To effectively utilize these transition-aware depth representations, we incorporate Structural Guidance Depth Modulator (SGDM), using a gated attention mechanism to efficiently fuse aligned depth and image features. Our proposed method achieves state-of-the-art performance on nuScenes validation dataset, with its mAP and NDS reaching 71.5% and 73.6% respectively.

View Paper

Pixels, Patterns, but No Poetry: To See The World like Humans

Authors: Hongcheng Gao, Zihao Huang, Lin Xu, Jingyi Tang, Xinhao Li, Yue Liu, Haoyang Li, Taihang Hu, Minhua Lin, Xinlong Yang, Ge Wu, Balong Bi, Hongyu Chen, Wentao Zhang

Abstract: arXiv:2507.16863v1 Announce Type: cross Abstract: Achieving human-like perception and reasoning in Multimodal Large Language Models (MLLMs) remains a central challenge in artificial intelligence. While recent research has primarily focused on enhancing reasoning capabilities in MLLMs, a fundamental question persists: Can Multimodal Large Language Models truly perceive the world as humans do? This paper shifts focus from reasoning to perception. Rather than constructing benchmarks specifically for reasoning, we introduce the Turing Eye Test (TET), a challenging perception-oriented benchmark comprising four diagnostic tasks that evaluate MLLMs' performance on synthetic images that humans process intuitively. Our findings reveal that state-of-the-art MLLMs exhibit catastrophic failures on our perceptual tasks trivial for humans. Both in-context learning and training on language backbone-effective for previous benchmarks-fail to improve performance on our tasks, while fine-tuning the vision tower enables rapid adaptation, suggesting that our benchmark poses challenges for vision tower generalization rather than for the knowledge and reasoning capabilities of the language backbone-a key gap between current MLLMs and human perception. We release a representative subset of TET tasks in this version, and will introduce more diverse tasks and methods to enhance visual generalization in future work.

View Paper

Reinforcement Learning in hyperbolic space for multi-step reasoning

Authors: Tao Xu, Dung-Yang Lee, Momiao Xiong

Abstract: arXiv:2507.16864v1 Announce Type: cross Abstract: Multi-step reasoning is a fundamental challenge in artificial intelligence, with applications ranging from mathematical problem-solving to decision-making in dynamic environments. Reinforcement Learning (RL) has shown promise in enabling agents to perform multi-step reasoning by optimizing long-term rewards. However, conventional RL methods struggle with complex reasoning tasks due to issues such as credit assignment, high-dimensional state representations, and stability concerns. Recent advancements in Transformer architectures and hyperbolic geometry have provided novel solutions to these challenges. This paper introduces a new framework that integrates hyperbolic Transformers into RL for multi-step reasoning. The proposed approach leverages hyperbolic embeddings to model hierarchical structures effectively. We present theoretical insights, algorithmic details, and experimental results that include Frontier Math and nonlinear optimal control problems. Compared to RL with vanilla transformer, the hyperbolic RL largely improves accuracy by (32%~44%) on FrontierMath benchmark, (43%~45%) on nonlinear optimal control benchmark, while achieving impressive reduction in computational time by (16%~32%) on FrontierMath benchmark, (16%~17%) on nonlinear optimal control benchmark. Our work demonstrates the potential of hyperbolic Transformers in reinforcement learning, particularly for multi-step reasoning tasks that involve hierarchical structures.

View Paper

Diffusion-Modeled Reinforcement Learning for Carbon and Risk-Aware Microgrid Optimization

Authors: Yunyi Zhao, Wei Zhang, Cheng Xiang, Hongyang Du, Dusit Niyato, Shuhua Gao

Abstract: arXiv:2507.16867v1 Announce Type: cross Abstract: This paper introduces DiffCarl, a diffusion-modeled carbon- and risk-aware reinforcement learning algorithm for intelligent operation of multi-microgrid systems. With the growing integration of renewables and increasing system complexity, microgrid communities face significant challenges in real-time energy scheduling and optimization under uncertainty. DiffCarl integrates a diffusion model into a deep reinforcement learning (DRL) framework to enable adaptive energy scheduling under uncertainty and explicitly account for carbon emissions and operational risk. By learning action distributions through a denoising generation process, DiffCarl enhances DRL policy expressiveness and enables carbon- and risk-aware scheduling in dynamic and uncertain microgrid environments. Extensive experimental studies demonstrate that it outperforms classic algorithms and state-of-the-art DRL solutions, with 2.3-30.1% lower operational cost. It also achieves 28.7% lower carbon emissions than those of its carbon-unaware variant and reduces performance variability. These results highlight DiffCarl as a practical and forward-looking solution. Its flexible design allows efficient adaptation to different system configurations and objectives to support real-world deployment in evolving energy systems.

View Paper

CompLeak: Deep Learning Model Compression Exacerbates Privacy Leakage

Authors: Na Li, Yansong Gao, Hongsheng Hu, Boyu Kuang, Anmin Fu

Abstract: arXiv:2507.16872v1 Announce Type: cross Abstract: Model compression is crucial for minimizing memory storage and accelerating inference in deep learning (DL) models, including recent foundation models like large language models (LLMs). Users can access different compressed model versions according to their resources and budget. However, while existing compression operations primarily focus on optimizing the trade-off between resource efficiency and model performance, the privacy risks introduced by compression remain overlooked and insufficiently understood. In this work, through the lens of membership inference attack (MIA), we propose CompLeak, the first privacy risk evaluation framework examining three widely used compression configurations that are pruning, quantization, and weight clustering supported by the commercial model compression framework of Google's TensorFlow-Lite (TF-Lite) and Facebook's PyTorch Mobile. CompLeak has three variants, given available access to the number of compressed models and original model. CompLeakNR starts by adopting existing MIA methods to attack a single compressed model, and identifies that different compressed models influence members and non-members differently. When the original model and one compressed model are available, CompLeakSR leverages the compressed model as a reference to the original model and uncovers more privacy by combining meta information (e.g., confidence vector) from both models. When multiple compressed models are available with/without accessing the original model, CompLeakMR innovatively exploits privacy leakage info from multiple compressed versions to substantially signify the overall privacy leakage. We conduct extensive experiments on seven diverse model architectures (from ResNet to foundation models of BERT and GPT-2), and six image and textual benchmark datasets.

View Paper

HIPPO-Video: Simulating Watch Histories with Large Language Models for Personalized Video Highlighting

Authors: Jeongeun Lee, Youngjae Yu, Dongha Lee

Abstract: arXiv:2507.16873v1 Announce Type: cross Abstract: The exponential growth of video content has made personalized video highlighting an essential task, as user preferences are highly variable and complex. Existing video datasets, however, often lack personalization, relying on isolated videos or simple text queries that fail to capture the intricacies of user behavior. In this work, we introduce HIPPO-Video, a novel dataset for personalized video highlighting, created using an LLM-based user simulator to generate realistic watch histories reflecting diverse user preferences. The dataset includes 2,040 (watch history, saliency score) pairs, covering 20,400 videos across 170 semantic categories. To validate our dataset, we propose HiPHer, a method that leverages these personalized watch histories to predict preference-conditioned segment-wise saliency scores. Through extensive experiments, we demonstrate that our method outperforms existing generic and query-based approaches, showcasing its potential for highly user-centric video highlighting in real-world scenarios.

View Paper

Budget Allocation Policies for Real-Time Multi-Agent Path Finding

Authors: Raz Beck, Roni Stern

Abstract: arXiv:2507.16874v1 Announce Type: cross Abstract: Multi-Agent Pathfinding (MAPF) is the problem of finding paths for a set of agents such that each agent reaches its desired destination while avoiding collisions with the other agents. Many MAPF solvers are designed to run offline, that is, first generate paths for all agents and then execute them. Real-Time MAPF (RT-MAPF) embodies a realistic MAPF setup in which one cannot wait until a complete path for each agent has been found before they start to move. Instead, planning and execution are interleaved, where the agents must commit to a fixed number of steps in a constant amount of computation time, referred to as the planning budget. Existing solutions to RT-MAPF iteratively call windowed versions of MAPF algorithms in every planning period, without explicitly considering the size of the planning budget. We address this gap and explore different policies for allocating the planning budget in windowed versions of standard MAPF algorithms, namely

Prioritized Planning (PrP) and MAPF-LNS2. Our exploration shows that the baseline approach in which all agents draw from a shared planning budget pool is ineffective in over-constrained situations. Instead, policies that distribute the planning budget over the agents are able to solve more problems with a smaller makespan.

View Paper

Machine learning-based multimodal prognostic models integrating pathology images and high-throughput omic data for overall survival prediction in cancer: a systematic review

Authors: Charlotte Jennings (National Pathology Imaging Cooperative, Leeds Teaching Hospitals NHS Trust, Leeds, UK), Andrew Broad (National Pathology Imaging Cooperative, Leeds Teaching Hospitals NHS Trust, Leeds, UK), Lucy Godson (National Pathology Imaging Cooperative, Leeds Teaching Hospitals NHS Trust, Leeds, UK), Emily Clarke (National Pathology Imaging Cooperative, Leeds Teaching Hospitals NHS Trust, Leeds, UK), David Westhead (University of Leeds, Leeds, UK), Darren Treanor (National Pathology Imaging Cooperative, Leeds Teaching Hospitals NHS Trust, Leeds, UK)

Abstract: arXiv:2507.16876v1 Announce Type: cross Abstract: Multimodal machine learning integrating histopathology and molecular data shows promise for cancer prognostication. We systematically reviewed studies combining whole slide images (WSIs) and high-throughput omics to predict overall survival. Searches of EMBASE, PubMed, and Cochrane CENTRAL (12/08/2024), plus citation screening, identified eligible studies. Data extraction used CHARMS; bias was assessed with PROBAST+AI; synthesis followed SWiM and PRISMA 2020. Protocol: PROSPERO (CRD42024594745). Forty-eight studies (all since 2017) across 19 cancer types met criteria; all used The Cancer Genome Atlas. Approaches included regularised Cox regression (n=4), classical ML (n=13), and deep learning (n=31). Reported c-indices ranged 0.550-0.857; multimodal models typically outperformed unimodal ones. However, all studies showed unclear/high bias, limited external validation, and little focus on clinical utility. Multimodal WSI-omics survival prediction is a fast-growing field with promising results but needs improved methodological rigor, broader datasets, and clinical evaluation. Funded by NPIC, Leeds Teaching Hospitals NHS Trust, UK (Project 104687), supported by UKRI Industrial Strategy Challenge Fund.

View Paper

ReMeREC: Relation-aware and Multi-entity Referring Expression Comprehension

Authors: Yizhi Hu, Zezhao Tian, Xingqun Qi, Chen Su, Bingkun Yang, Junhui Yin, Muyi Sun, Man Zhang, Zhenan Sun

Abstract: arXiv:2507.16877v1 Announce Type: cross Abstract: Referring Expression Comprehension (REC) aims to localize specified entities or regions in an image based on natural language descriptions. While existing methods handle single-entity localization, they often ignore complex inter-entity relationships in multi-entity scenes, limiting their accuracy and reliability. Additionally, the lack of high-quality datasets with fine-grained, paired image-text-relation annotations hinders further progress. To address this challenge, we first construct a relation-aware, multi-entity REC dataset called ReMeX, which includes detailed relationship and textual annotations. We then propose ReMeREC, a novel framework that jointly leverages visual and textual cues to localize multiple entities while modeling their inter-relations. To address the semantic ambiguity caused by implicit entity boundaries in language, we introduce the Text-adaptive Multi-entity Perceptron (TMP), which dynamically infers both the quantity and span of entities from fine-grained textual cues, producing distinctive representations. Additionally, our Entity Inter-relationship Reasoner (EIR) enhances relational reasoning and global scene understanding. To further improve language comprehension for fine-grained prompts, we also construct a small-scale auxiliary dataset, EntityText, generated using large language models. Experiments on four benchmark datasets show that ReMeREC achieves state-of-the-art performance in multi-entity grounding and relation prediction, outperforming existing approaches by a large margin.

View Paper

CausalStep: A Benchmark for Explicit Stepwise Causal Reasoning in Videos

Authors: Xuchen Li, Xuzhao Li, Shiyu Hu, Kaiqi Huang, Wentao Zhang

Abstract: arXiv:2507.16878v1 Announce Type: cross Abstract: Recent advances in large language models (LLMs) have improved reasoning in text and image domains, yet achieving robust video reasoning remains a significant challenge. Existing video benchmarks mainly assess shallow understanding and reasoning and allow models to exploit global context, failing to rigorously evaluate true causal and stepwise reasoning. We present CausalStep, a benchmark designed for explicit stepwise causal reasoning in videos. CausalStep segments videos into causally linked units and enforces a strict stepwise question-answer (QA)

protocol, requiring sequential answers and preventing shortcut solutions. Each question includes carefully constructed distractors based on error type taxonomy to ensure diagnostic value. The benchmark features 100 videos across six categories and 1,852 multiple-choice QA pairs. We introduce seven diagnostic metrics for comprehensive evaluation, enabling precise diagnosis of causal reasoning capabilities. Experiments with leading proprietary and open-source models, as well as human baselines, reveal a significant gap between current models and human-level stepwise reasoning. CausalStep provides a rigorous benchmark to drive progress in robust and interpretable video reasoning.

View Paper

Finding Dori: Memorization in Text-to-Image Diffusion Models Is Less Local Than Assumed

Authors: Antoni Kowalczuk, Dominik Hintersdorf, Lukas Struppek, Kristian Kersting, Adam Dziedzic, Franziska Boenisch

Abstract: arXiv:2507.16880v1 Announce Type: cross Abstract: Text-to-image diffusion models (DMs) have achieved remarkable success in image generation. However, concerns about data privacy and intellectual property remain due to their potential to inadvertently memorize and replicate training data. Recent mitigation efforts have focused on identifying and pruning weights responsible for triggering replication, based on the assumption that memorization can be localized. Our research assesses the robustness of these pruning-based approaches. We demonstrate that even after pruning, minor adjustments to text embeddings of input prompts are sufficient to re-trigger data replication, highlighting the fragility of these defenses. Furthermore, we challenge the fundamental assumption of memorization locality, by showing that replication can be triggered from diverse locations within the text embedding space, and follows different paths in the model. Our findings indicate that existing mitigation strategies are insufficient and underscore the need for methods that truly remove memorized content, rather than attempting to suppress its retrieval. As a first step in this direction, we introduce a novel adversarial fine-tuning method that iteratively searches for replication triggers and updates the model to increase robustness. Through our research, we provide fresh insights into the nature of memorization in text-to-image DMs and a foundation for building more trustworthy and compliant generative AI.

View Paper

Confidence Optimization for Probabilistic Encoding

Authors: Pengjiu Xia, Yidian Huang, Wenchao Wei, Yuwen Tan

Abstract: arXiv:2507.16881v1 Announce Type: cross Abstract: Probabilistic encoding introduces Gaussian noise into neural networks, enabling a smooth transition from deterministic to uncertain states and enhancing generalization ability. However, the randomness of Gaussian noise distorts point-based distance measurements in classification tasks. To mitigate this issue, we propose a confidence optimization probabilistic encoding (CPE) method that improves distance reliability and enhances representation learning. Specifically, we refine probabilistic encoding with two key strategies: First, we introduce a confidence-aware mechanism to adjust distance calculations, ensuring consistency and reliability in probabilistic encoding classification tasks. Second, we replace the conventional KL divergence-based variance regularization, which relies on unreliable prior assumptions, with a simpler L2 regularization term to directly constrain variance. The method we proposed is model-agnostic, and extensive experiments on natural language classification tasks demonstrate that our method significantly improves performance and generalization on both the BERT and the RoBERTa model.

View Paper

SplitMeanFlow: Interval Splitting Consistency in Few-Step Generative Modeling

Authors: Yi Guo, Wei Wang, Zhihang Yuan, Rong Cao, Kuan Chen, Zhengyang Chen, Yuanyuan Huo, Yang Zhang, Yuping Wang, Shouda Liu, Yuxuan Wang

Abstract: arXiv:2507.16884v1 Announce Type: cross Abstract: Generative models like Flow Matching have achieved state-of-the-art performance but are often hindered by a computationally expensive iterative sampling process. To address this, recent work has focused on few-step or one-step generation by learning the average velocity field, which directly maps noise to data. MeanFlow, a leading method in this area, learns this field by enforcing a differential identity that connects the average and instantaneous velocities. In this work, we argue that this differential formulation is a limiting special case of a more fundamental principle. We return to the first principles of average velocity and leverage the additivity property of definite integrals. This leads us to derive a novel, purely algebraic identity we term Interval Splitting Consistency. This identity establishes a self-referential relationship for the average velocity field across different time intervals without resorting to any differential operators. Based on this principle, we introduce SplitMeanFlow, a new training framework that enforces this algebraic consistency directly as a learning objective. We formally prove that the differential identity at the core of MeanFlow is recovered by taking the limit of our algebraic consistency as the interval split becomes infinitesimal. This establishes SplitMeanFlow as a direct and more general foundation for learning average velocity fields. From a practical standpoint, our algebraic approach is

significantly more efficient, as it eliminates the need for JVP computations, resulting in simpler implementation, more stable training, and broader hardware compatibility. One-step and two-step SplitMeanFlow models have been successfully deployed in large-scale speech synthesis products (such as Doubao), achieving speedups of 20x.

View Paper

Sparser2Sparse: Single-shot Sparser-to-Sparse Learning for Spatial Transcriptomics Imputation with Natural Image Co-learning

Authors: Yaoyu Fang, Jiahe Qian, Xinkun Wang, Lee A. Cooper, Bo Zhou

Abstract: arXiv:2507.16886v1 Announce Type: cross Abstract: Spatial transcriptomics (ST) has revolutionized biomedical research by enabling high resolution gene expression profiling within tissues. However, the high cost and scarcity of high resolution ST data remain significant challenges. We present Single-shot Sparser-to-Sparse (S2S-ST), a novel framework for accurate ST imputation that requires only a single and low-cost sparsely sampled ST dataset alongside widely available natural images for co-training. Our approach integrates three key innovations: (1) a sparser-to-sparse self-supervised learning strategy that leverages intrinsic spatial patterns in ST data, (2) cross-domain co-learning with natural images to enhance feature representation, and (3) a Cascaded Data Consistent Imputation Network (CDCIN) that iteratively refines predictions while preserving sampled gene data fidelity. Extensive experiments on diverse tissue types, including breast cancer, liver, and lymphoid tissue, demonstrate that our method outperforms state-of-the-art approaches in imputation accuracy. By enabling robust ST reconstruction from sparse inputs, our framework significantly reduces reliance on costly high resolution data, facilitating potential broader adoption in biomedical research and clinical applications.

View Paper

Revisiting Pre-trained Language Models for Vulnerability Detection

Authors: Youpeng Li, Weiliang Qi, Xuyu Wang, Fuxun Yu, Xinda Wang

Abstract: arXiv:2507.16887v1 Announce Type: cross Abstract: The rapid advancement of pre-trained language models (PLMs) has demonstrated promising results for various code-related tasks. However, their effectiveness in detecting real-world vulnerabilities remains a critical challenge. % for the security community. While existing empirical studies evaluate PLMs for vulnerability detection (VD), their inadequate consideration in data preparation, evaluation setups, and experimental settings undermines the accuracy and comprehensiveness of evaluations. This paper introduces RevisitVD, an extensive evaluation of 17 PLMs spanning smaller code-specific PLMs and large-scale PLMs using newly constructed datasets. Specifically, we compare the performance of PLMs under both fine-tuning and prompt engineering, assess their effectiveness and generalizability across various training and testing settings, and analyze their robustness against code normalization, abstraction, and semantic-preserving transformations. Our findings reveal that, for VD tasks, PLMs incorporating pre-training tasks designed to capture the syntactic and semantic patterns of code outperform both general-purpose PLMs and those solely pre-trained or fine-tuned on large code corpora. However, these models face notable challenges in real-world scenarios, such as difficulties in detecting vulnerabilities with complex dependencies, handling perturbations introduced by code normalization and abstraction, and identifying semantic-preserving vulnerable code transformations. Also, the truncation caused by the limited context windows of PLMs can lead to a non-negligible amount of labeling errors. This study underscores the importance of thorough evaluations of model performance in practical scenarios and outlines future directions to help enhance the effectiveness of PLMs for realistic VD applications.

View Paper

SiLQ: Simple Large Language Model Quantization-Aware Training

Authors: Steven K. Esser, Jeffrey L. McKinstry, Deepika Bablani, Rathinakumar Appuswamy, Dharmendra S. Modha

Abstract: arXiv:2507.16933v1 Announce Type: cross Abstract: Large language models can be quantized to reduce inference time latency, model size, and energy consumption, thereby delivering a better user experience at lower cost. A challenge exists to deliver quantized models with minimal loss of accuracy in reasonable time, and in particular to do so without requiring mechanisms incompatible with specialized inference accelerators. Here, we demonstrate a simple, end-to-end quantization-aware training approach that, with an increase in total model training budget of less than 0.1%, outperforms the leading published quantization methods by large margins on several modern benchmarks, with both base and instruct model variants. The approach easily generalizes across different model architectures, can be applied to activations, cache, and weights, and requires the introduction of no additional operations to the model other than the quantization itself.

View Paper

Evaluating Ensemble and Deep Learning Models for Static Malware Detection with Dimensionality Reduction Using the EMBER Dataset

Authors: Md Min-Ha-Zul Abedin, Tazqia Mehrub

Abstract: arXiv:2507.16952v1 Announce Type: cross Abstract: This study investigates the effectiveness of several machine learning algorithms for static malware detection using the EMBER dataset, which contains feature representations of Portable Executable (PE) files. We evaluate eight classification models: LightGBM, XGBoost, CatBoost, Random Forest, Extra Trees, HistGradientBoosting, k-Nearest Neighbors (KNN), and TabNet, under three preprocessing settings: original feature space, Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA). The models are assessed on accuracy, precision, recall, F1 score, and AUC to examine both predictive performance and robustness. Ensemble methods, especially LightGBM and XGBoost, show the best overall performance across all configurations, with minimal sensitivity to PCA and consistent generalization. LDA improves KNN performance but significantly reduces accuracy for boosting models. TabNet, while promising in theory, underperformed under feature reduction, likely due to architectural sensitivity to input structure. The analysis is supported by detailed exploratory data analysis (EDA), including mutual information ranking, PCA or t-SNE visualizations, and outlier detection using Isolation Forest and Local Outlier Factor (LOF), which confirm the discriminatory capacity of key features in the EMBER dataset. The results suggest that boosting models remain the most reliable choice for high-dimensional static malware detection, and that dimensionality reduction should be applied selectively based on model type. This work provides a benchmark for comparing classification models and preprocessing strategies in malware detection tasks and contributes insights that can guide future system development and real-world deployment.

View Paper

Text-to-SPARQL Goes Beyond English: Multilingual Question Answering Over Knowledge Graphs through Human-Inspired Reasoning

Authors: Aleksandr Perevalov, Andreas Both

Abstract: arXiv:2507.16971v1 Announce Type: cross Abstract: Accessing knowledge via multilingual natural-language interfaces is one of the emerging challenges in the field of information retrieval and related ones. Structured knowledge stored in knowledge graphs can be queried via a specific query language (e.g., SPARQL). Therefore, one needs to transform natural-language input into a query to fulfill an information need. Prior approaches mostly focused on combining components (e.g., rule-based or neural-based) that solve downstream tasks and come up with an answer at the end. We introduce mKGQAgent, a human-inspired framework that breaks down the task of converting natural language questions into SPARQL queries into modular, interpretable subtasks. By leveraging a coordinated LLM agent workflow for planning, entity linking, and query refinement - guided by an experience pool for in-context learning - mKGQAgent efficiently handles multilingual KGQA. Evaluated on the DBpedia- and Corporate-based KGQA benchmarks within the Text2SPARQL challenge 2025, our approach took first place among the other participants. This work opens new avenues for developing human-like reasoning systems in multilingual semantic parsing.

View Paper

Leveraging Synthetic Data for Question Answering with Multilingual LLMs in the Agricultural Domain

Authors: Rishemjit Kaur, Arshdeep Singh Bhankhar, Surangika Ranathunga, Jashanpreet Singh Salh, Sudhir Rajput, Vidhi, Kashish Mahendra, Bhavika Berwal, Ritesh Kumar

Abstract: arXiv:2507.16974v1 Announce Type: cross Abstract: Enabling farmers to access accurate agriculture-related information in their native languages in a timely manner is crucial for the success of the agriculture field. Although large language models (LLMs) can be used to implement Question Answering (QA) systems, simply using publicly available general-purpose LLMs in agriculture typically offer generic advisories, lacking precision in local and multilingual contexts due to insufficient domain-specific training and scarcity of high-quality, region-specific datasets. Our study addresses these limitations by generating multilingual synthetic agricultural datasets (English, Hindi, Punjabi) from agriculture-specific documents and fine-tuning language-specific LLMs. Our evaluation on curated multilingual datasets demonstrates significant improvements in factual accuracy, relevance, and agricultural consensus for the fine-tuned models compared to their baseline counterparts. These results highlight the efficacy of synthetic data-driven, language-specific fine-tuning as an effective strategy to improve the performance of LLMs in agriculture, especially in multilingual and low-resource settings. By enabling more accurate and localized agricultural advisory services, this study provides a meaningful step

toward bridging the knowledge gap in AI-driven agricultural solutions for diverse linguistic communities.

View Paper

Fast and Scalable Gene Embedding Search: A Comparative Study of FAISS and ScaNN

Authors: Mohammad Saleh Refahi, Gavin Hearne, Harrison Muller, Kieran Lynch, Bahrad A. Sokhansanj, James R. Brown, Gail Rosen

Abstract: arXiv:2507.16978v1 Announce Type: cross Abstract: The exponential growth of DNA sequencing data has outpaced traditional heuristic-based methods, which struggle to scale effectively. Efficient computational approaches are urgently needed to support large-scale similarity search, a foundational task in bioinformatics for detecting homology, functional similarity, and novelty among genomic and proteomic sequences. Although tools like BLAST have been widely used and remain effective in many scenarios, they suffer from limitations such as high computational cost and poor performance on divergent sequences. In this work, we explore embedding-based similarity search methods that learn latent representations capturing deeper structural and functional patterns beyond raw sequence alignment. We systematically evaluate two state-of-the-art vector search libraries, FAISS and ScaNN, on biologically meaningful gene embeddings. Unlike prior studies, our analysis focuses on bioinformatics-specific embeddings and benchmarks their utility for detecting novel sequences, including those from uncharacterized taxa or genes lacking known homologs. Our results highlight both computational advantages (in memory and runtime efficiency) and improved retrieval quality, offering a promising alternative to traditional alignment-heavy tools.

View Paper

PyG 2.0: Scalable Learning on Real World Graphs

Authors: Matthias Fey, Jinu Sunil, Akihiro Nitta, Rishi Puri, Manan Shah, Bla\v{z} Stojanovi\v{c}, Ramona Bendias, Alexandria Barghi, Vid Kocijan, Zecheng Zhang, Xinwei He, Jan Eric Lenssen, Jure Leskovec

Abstract: arXiv:2507.16991v1 Announce Type: cross Abstract: PyG (PyTorch Geometric) has evolved significantly since its initial release, establishing itself as a leading framework for Graph Neural Networks. In this paper, we present Pyg 2.0 (and its subsequent minor versions), a comprehensive update that introduces substantial improvements in scalability and real-world application capabilities. We detail the framework's enhanced architecture, including support for heterogeneous and temporal graphs, scalable feature/graph stores, and various optimizations, enabling researchers and practitioners to tackle large-scale graph learning problems efficiently. Over the recent years, PyG has been supporting graph learning in a large variety of application areas, which we will summarize, while providing a deep dive into the important areas of relational deep learning and large language modeling.

View Paper

Bayesian preference elicitation for decision support in multiobjective optimization

Authors: Felix Huber, Sebastian Rojas Gonzalez, Raul Astudillo

Abstract: arXiv:2507.16999v1 Announce Type: cross Abstract: We present a novel approach to help decision-makers efficiently identify preferred solutions from the Pareto set of a multi-objective optimization problem. Our method uses a Bayesian model to estimate the decision-maker's utility function based on pairwise comparisons. Aided by this model, a principled elicitation strategy selects queries interactively to balance exploration and exploitation, guiding the discovery of high-utility solutions. The approach is flexible: it can be used interactively or a posteriori after estimating the Pareto front through standard multi-objective optimization techniques. Additionally, at the end of the elicitation phase, it generates a reduced menu of high-quality solutions, simplifying the decision-making process. Through experiments on test problems with up to nine objectives, our method demonstrates superior performance in finding high-utility solutions with a small number of queries. We also provide an open-source implementation of our method to support its adoption by the broader community.

View Paper

Bringing Balance to Hand Shape Classification: Mitigating Data Imbalance Through Generative Models

Authors: Gaston Gustavo Rios, Pedro Dal Bianco, Franco Ronchetti, Facundo Quiroga, Oscar Stanchi, Santiago Ponte Ah\'on, Waldo Hasperu\'e

Abstract: arXiv:2507.17008v1 Announce Type: cross Abstract: Most sign language handshape datasets are severely limited and unbalanced, posing significant challenges to effective model training. In this paper, we explore the effectiveness of augmenting the training data of a handshape classifier by generating synthetic data. We use an EfficientNet classifier trained on the RWTH German sign language handshape dataset, which is small and heavily unbalanced, applying different strategies to combine generated and real images. We compare two Generative Adversarial Networks (GAN) architectures for data generation: ReACGAN, which uses label information to condition the data generation process through an auxiliary classifier, and SPADE, which utilizes spatially-adaptive normalization to condition the generation on pose information. ReACGAN allows for the generation of realistic images that align with specific handshape labels, while SPADE focuses on generating images with accurate spatial handshape configurations. Our proposed techniques improve the current state-of-the-art accuracy on the RWTH dataset by 5%, addressing the limitations of small and unbalanced datasets. Additionally, our method demonstrates the capability to generalize across different sign language datasets by leveraging pose-based generation trained on the extensive HaGRID dataset. We achieve comparable performance to single-source trained classifiers without the need for retraining the generator.

View Paper

Towards Trustworthy AI: Secure Deepfake Detection using CNNs and Zero-Knowledge Proofs

Authors: H M Mohaimanul Islam, Huynh Q. N. Vo, Aditya Rane

Abstract: arXiv:2507.17010v1 Announce Type: cross Abstract: In the era of synthetic media, deepfake manipulations pose a significant threat to information integrity. To address this challenge, we propose TrustDefender, a two-stage framework comprising (i) a lightweight convolutional neural network (CNN) that detects deepfake imagery in real-time extended reality (XR) streams, and (ii) an integrated succinct zero-knowledge proof (ZKP) protocol that validates detection results without disclosing raw user data. Our design addresses both the computational constraints of XR platforms while adhering to the stringent privacy requirements in sensitive settings. Experimental evaluations on multiple benchmark deepfake datasets demonstrate that TrustDefender achieves 95.3% detection accuracy, coupled with efficient proof generation underpinned by rigorous cryptography, ensuring seamless integration with high-performance artificial intelligence (AI) systems. By fusing advanced computer vision models with provable security mechanisms, our work establishes a foundation for reliable AI in immersive and privacy-sensitive applications.

View Paper

laplax -- Laplace Approximations with JAX

Authors: Tobias Weber, B\'alint Mucs\'anyi, Lenard Rommel, Thomas Christie, Lars Kas\"uschke, Marvin Pf\"ortner, Philipp Hennig

Abstract: arXiv:2507.17013v1 Announce Type: cross Abstract: The Laplace approximation provides a scalable and efficient means of quantifying weight-space uncertainty in deep neural networks, enabling the application of Bayesian tools such as predictive uncertainty and model selection via Occam's razor. In this work, we introduce laplax, a new open-source Python package for performing Laplace approximations with jax. Designed with a modular and purely functional architecture and minimal external dependencies, laplax offers a flexible and researcher-friendly framework for rapid prototyping and experimentation. Its goal is to facilitate research on Bayesian neural networks, uncertainty quantification for deep learning, and the development of improved Laplace approximation techniques.

View Paper

Can External Validation Tools Improve Annotation Quality for LLM-as-a-Judge?

Authors: Arduin Findeis, Floris Weers, Guoli Yin, Ke Ye, Ruoming Pang, Tom Gunter

Abstract: arXiv:2507.17015v1 Announce Type: cross Abstract: Pairwise preferences over model responses are widely collected to evaluate and provide feedback to large language models (LLMs). Given two alternative model responses to the same input, a human or AI annotator selects the "better" response. This approach can provide feedback for domains where other hard-coded metrics are difficult to obtain (e.g., chat response quality), thereby helping model evaluation or training. However, for some domains high-quality pairwise comparisons can be tricky to obtain - from AI and humans. For example, for responses with many factual statements, annotators may disproportionately weigh writing quality rather than underlying facts. In this work, we explore augmenting standard AI annotator systems with additional tools to improve performance on three challenging response domains: long-form factual, math and code tasks. We propose a tool-using agentic system to provide higher quality feedback on these domains. Our system uses web-search and code execution to ground itself based on external validation, independent of the LLM's internal knowledge and biases. We provide extensive experimental results evaluating our method across the three targeted response domains as well as general annotation tasks, using RewardBench (incl. AlpacaEval and LLMBar), RewardMath, as well as three new datasets for domains with saturated pre-existing datasets. Our results indicate that external tools can indeed improve performance in many, but not all, cases. More generally, our experiments highlight the sensitivity of performance to simple parameters (e.g., prompt) and the need for improved (non-saturated) annotator benchmarks. We share our code at https://github.com/apple/ml-agent-evaluator.

View Paper

Causal Graph Fuzzy LLMs: A First Introduction and Applications in Time Series Forecasting

Authors: Omid Orang, Patricia O. Lucas, Gabriel I. F. Paiva, Petronio C. L. Silva, Felipe Augusto Rocha da Silva, Adriano Alonso Veloso, Frederico Gadelha Guimaraes

Abstract: arXiv:2507.17016v1 Announce Type: cross Abstract: In recent years, the application of Large Language Models (LLMs) to time series forecasting (TSF) has garnered significant attention among researchers. This study presents a new frame of LLMs named CGF-LLM using GPT-2 combined with fuzzy time series (FTS) and causal graph to predict multivariate time series, marking the first such architecture in the literature. The key objective is to convert numerical time series into interpretable forms through the parallel application of fuzzification and causal analysis, enabling both semantic understanding and structural insight as input for the pretrained GPT-2 model. The resulting textual representation offers a more interpretable view of the complex dynamics underlying the original time series. The reported results confirm the effectiveness of our proposed LLM-based time series forecasting model, as demonstrated across four different multivariate time series datasets. This initiative paves promising future directions in the domain of TSF using LLMs based on FTS.

View Paper

Evolutionary Feature-wise Thresholding for Binary Representation of NLP Embeddings

Authors: Soumen Sinha, Shahryar Rahnamayan, Azam Asilian Bidgoli

Abstract: arXiv:2507.17025v1 Announce Type: cross Abstract: Efficient text embedding is crucial for large-scale natural language processing (NLP) applications, where storage and computational efficiency are key concerns. In this paper, we explore how using binary representations (barcodes) instead of real-valued features can be used for NLP embeddings derived from machine learning models such as BERT. Thresholding is a common method for converting continuous embeddings into binary representations, often using a fixed threshold across all features. We propose a Coordinate Search-based optimization framework that instead identifies the optimal threshold for each feature, demonstrating that feature-specific thresholds lead to improved performance in binary encoding. This ensures that the binary representations are both accurate and efficient, enhancing performance across various features. Our optimal barcode representations have shown promising results in various NLP applications, demonstrating their potential to transform text representation. We conducted extensive experiments and statistical tests on different NLP tasks and datasets to evaluate our approach and compare it to other thresholding methods. Binary embeddings generated using using optimal thresholds found by our method outperform traditional binarization methods in accuracy. This technique for generating binary representations is versatile and can be applied to any features, not just limited to NLP embeddings, making it useful for a wide range of domains in machine learning applications.

View Paper

StreamME: Simplify 3D Gaussian Avatar within Live Stream

Authors: Luchuan Song, Yang Zhou, Zhan Xu, Yi Zhou, Deepali Aneja, Chenliang Xu

Abstract: arXiv:2507.17029v1 Announce Type: cross Abstract: We propose StreamME, a method focuses on fast 3D avatar reconstruction. The StreamME synchronously records and reconstructs a head avatar from live video streams without any pre-cached data, enabling seamless integration of the reconstructed appearance into downstream applications. This exceptionally fast training strategy, which we refer to as on-the-fly training, is central to our approach. Our method is built upon 3D Gaussian Splatting (3DGS), eliminating the reliance on MLPs in deformable 3DGS and relying solely on geometry, which significantly improves the adaptation speed to facial expression. To further ensure high efficiency in on-the-fly training, we introduced a simplification strategy based on primary points, which distributes the point clouds more sparsely across the facial surface, optimizing points number while maintaining rendering quality. Leveraging the on-the-fly training capabilities, our method protects the facial privacy and reduces communication bandwidth in VR system or online conference. Additionally, it can be directly applied to downstream application such as animation, toonify, and relighting. Please refer to our project page for more details: https://songluchuan.github.io/StreamME/.

View Paper

Computational Performance Bounds Prediction in Quantum Computing with Unstable Noise

Authors: Jinyang Li, Samudra Dasgupta, Yuhong Song, Lei Yang, Travis Humble, Weiwen Jiang

Abstract: arXiv:2507.17043v1 Announce Type: cross Abstract: Quantum computing has significantly advanced in recent years, boasting devices with hundreds of quantum bits (qubits), hinting at its potential quantum advantage over classical computing. Yet, noise in quantum devices poses significant barriers to realizing this supremacy. Understanding noise's impact is crucial for reproducibility and application reuse; moreover, the next-generation quantum-centric supercomputing essentially requires efficient and accurate noise characterization to support system management (e.g., job scheduling), where ensuring correct functional performance (i.e., fidelity) of jobs on available quantum devices can even be higher-priority than traditional objectives. However, noise fluctuates over time, even on the same quantum device, which makes predicting the computational bounds for on-the-fly noise is vital. Noisy quantum simulation can offer insights but faces efficiency and scalability issues. In this work, we propose a data-driven workflow, namely QuBound, to predict computational performance bounds. It decomposes historical performance traces to isolate noise sources and devises a novel encoder to embed circuit and noise information processed by a Long Short-Term Memory (LSTM) network. For evaluation, we compare QuBound with a state-of-the-art learning-based predictor, which only generates a single performance value instead of a bound. Experimental results show that the result of the existing approach falls outside of performance bounds, while all predictions from our QuBound with the assistance of performance decomposition better fit the bounds. Moreover, QuBound can efficiently produce practical bounds for various circuits with over 106 speedup over simulation; in addition, the range from QuBound is over 10x narrower than the state-of-the-art analytical approach.

View Paper

Controllable Hybrid Captioner for Improved Long-form Video Understanding

Authors: Kuleen Sasse, Efsun Sarioglu Kayi, Arun Reddy

Abstract: arXiv:2507.17047v1 Announce Type: cross Abstract: Video data, especially long-form video, is extremely dense and high-dimensional. Text-based summaries of video content offer a way to represent query-relevant content in a much more compact manner than raw video. In addition, textual representations are easily ingested by state-of-the-art large language models (LLMs), which enable reasoning over video content to answer complex natural language queries. To solve this issue, we rely on the progressive construction of a text-based memory by a video captioner operating on shorter chunks of the video, where spatio-temporal modeling is computationally feasible. We explore ways to improve the quality of the activity log comprised solely of short video captions. Because the video captions tend to be focused on human actions, and questions may pertain to other information in the scene, we seek to enrich the memory with static scene descriptions using Vision Language Models (VLMs). Our video understanding system relies on the LaViLa video captioner in combination with a LLM to answer questions about videos. We first explored different ways of partitioning the video into meaningful segments such that the textual descriptions more accurately reflect the structure of the video content. Furthermore, we incorporated static scene descriptions into the captioning pipeline using LLaVA VLM, resulting in a more detailed and complete caption log and expanding the space of questions that are answerable from the textual memory. Finally, we have successfully fine-tuned the LaViLa video captioner to produce both action and scene captions, significantly improving the efficiency of the captioning pipeline compared to using separate captioning models for the two tasks. Our model, controllable hybrid captioner, can alternate between different types of captions according to special input tokens that signals scene changes detected in the video.

View Paper

Pragmatic Policy Development via Interpretable Behavior Cloning

Authors: Anton Matsson, Yaochen Rao, Heather J. Litman, Fredrik D. Johansson

Abstract: arXiv:2507.17056v1 Announce Type: cross Abstract: Offline reinforcement learning (RL) holds great promise for deriving optimal policies from observational data, but challenges related to interpretability and evaluation limit its practical use in safety-critical domains. Interpretability is hindered by the black-box nature of unconstrained RL policies, while evaluation -- typically performed off-policy -- is sensitive to large deviations from the data-collecting behavior policy, especially when using methods based on importance sampling. To address these challenges, we propose a simple yet practical alternative: deriving treatment policies from the most frequently chosen actions in each patient state, as estimated by an interpretable model of the behavior policy. By using a tree-based model, which is specifically designed to exploit patterns in the data, we obtain a natural grouping of states with respect to treatment. The tree structure ensures interpretability by design, while varying the number of actions considered controls the degree of overlap with the behavior policy, enabling reliable off-policy evaluation. This pragmatic approach to policy development standardizes frequent treatment patterns, capturing the collective clinical judgment embedded in the data. Using real-world examples in rheumatoid arthritis and sepsis care, we demonstrate that policies derived under this framework can outperform current practice, offering interpretable alternatives to those obtained via offline RL.

View Paper

Parallelism Meets Adaptiveness: Scalable Documents Understanding in Multi-Agent LLM Systems

Authors: Chengxuan Xia, Qianye Wu, Sixuan Tian, Yilun Hao

Abstract: arXiv:2507.17061v1 Announce Type: cross Abstract: Large language model (LLM) agents have shown increasing promise for collaborative task completion. However, existing multi-agent frameworks often rely on static workflows, fixed roles, and limited inter-agent communication, reducing their effectiveness in open-ended, high-complexity domains. This paper proposes a coordination framework that enables adaptiveness through three core mechanisms: dynamic task routing, bidirectional feedback, and parallel agent evaluation. The framework allows agents to reallocate tasks based on confidence and workload, exchange structured critiques to iteratively improve outputs, and crucially compete on high-ambiguity subtasks with evaluator-driven selection of the most suitable result. We instantiate these principles in a modular architecture and demonstrate substantial improvements in factual coverage, coherence, and efficiency over static and partially adaptive baselines. Our findings highlight the benefits of incorporating both adaptiveness and structured competition in multi-agent LLM systems.

View Paper

Compatibility of Max and Sum Objectives for Committee Selection and $k$-Facility Location

Authors: Yue Han, Elliot Anshelevich

Abstract: arXiv:2507.17063v1 Announce Type: cross Abstract: We study a version of the metric facility location problem (or, equivalently, variants of the committee selection problem) in which we must choose $k$ facilities in an arbitrary metric space to serve some set of clients $C$. We consider four different objectives, where each client $i\in C$ attempts to minimize either the sum or the maximum of its distance to the chosen facilities, and where the overall objective either considers the sum or the maximum of the individual client costs. Rather than optimizing a single objective at a time, we study how compatible these objectives are with each other, and show the existence of solutions which are simultaneously close-to-optimum for any pair of the above objectives. Our results show that when choosing a set of facilities or a representative committee, it is often possible to form a solution which is good for several objectives at the same time, instead of sacrificing one desideratum to achieve another.

View Paper

Advancing Robustness in Deep Reinforcement Learning with an Ensemble Defense Approach

Authors: Adithya Mohan, Dominik R\"o{\ss}le, Daniel Cremers, Torsten Sch\"on

Abstract: arXiv:2507.17070v1 Announce Type: cross Abstract: Recent advancements in Deep Reinforcement Learning (DRL) have demonstrated its applicability across various domains, including robotics, healthcare, energy optimization, and autonomous driving. However, a critical question remains: How robust are DRL models when exposed to adversarial attacks? While existing defense mechanisms such as adversarial training and distillation enhance the resilience of DRL models, there remains a significant research gap regarding the integration of multiple defenses in autonomous driving scenarios specifically. This paper addresses this gap by proposing a novel ensemble-based defense architecture to mitigate adversarial attacks in autonomous driving. Our evaluation demonstrates that the proposed architecture significantly enhances the robustness of DRL models. Compared to the baseline under FGSM attacks, our ensemble method improves the mean reward from 5.87 to 18.38 (over 213% increase) and reduces the mean collision rate from 0.50 to 0.09 (an 82% decrease) in the highway scenario and merge scenario, outperforming all standalone defense strategies.

View Paper

VL-CLIP: Enhancing Multimodal Recommendations via Visual Grounding and LLM-Augmented CLIP Embeddings

Authors: Ramin Giahi, Kehui Yao, Sriram Kollipara, Kai Zhao, Vahid Mirjalili, Jianpeng Xu, Topojoy Biswas, Evren Korpeoglu, Kannan Achan

Abstract: arXiv:2507.17080v1 Announce Type: cross Abstract: Multimodal learning plays a critical role in e-commerce recommendation platforms today, enabling accurate recommendations and product understanding. However, existing vision-language models, such as CLIP, face key challenges in e-commerce recommendation systems: 1) Weak object-level alignment, where global image embeddings fail to capture fine-grained product attributes, leading to suboptimal retrieval performance; 2) Ambiguous textual representations, where product descriptions often lack contextual clarity, affecting cross-modal matching; and 3) Domain mismatch, as generic vision-language models may not generalize well to e-commerce-specific data. To address these limitations, we propose a framework, VL-CLIP, that enhances CLIP embeddings by integrating Visual Grounding for fine-grained visual understanding and an LLM-based agent for generating enriched text embeddings. Visual Grounding refines image representations by localizing key products, while the LLM agent enhances textual features by disambiguating product descriptions. Our approach significantly improves retrieval accuracy, multimodal retrieval effectiveness, and recommendation quality across tens of millions of items on one of the largest e-commerce platforms in the U.S., increasing CTR by 18.6%, ATC by 15.5%, and GMV by 4.0%. Additional experimental results show that our framework outperforms vision-language models, including CLIP, FashionCLIP, and GCL, in both precision and semantic alignment, demonstrating the potential of combining object-aware visual grounding and LLM-enhanced text representation for robust multimodal recommendations.

View Paper

SDGOCC: Semantic and Depth-Guided Bird's-Eye View Transformation for 3D Multimodal Occupancy Prediction

Authors: Zaipeng Duan, Chenxu Dang, Xuzhong Hu, Pei An, Junfeng Ding, Jie Zhan, Yunbiao Xu, Jie Ma

Abstract: arXiv:2507.17083v1 Announce Type: cross Abstract: Multimodal 3D occupancy prediction has garnered significant attention for its potential in autonomous driving. However, most existing approaches are single-modality: camera-based methods lack depth information, while LiDAR-based methods struggle with occlusions. Current lightweight methods primarily rely on the Lift-Splat-Shoot (LSS) pipeline, which suffers from inaccurate depth estimation and fails to fully exploit the geometric and semantic information of 3D LiDAR points. Therefore, we propose a novel multimodal occupancy prediction network called SDG-OCC, which incorporates a joint semantic and depth-guided view transformation coupled with a fusion-to-occupancy-driven active distillation. The

enhanced view transformation constructs accurate depth distributions by integrating pixel semantics and co-point depth through diffusion and bilinear discretization. The fusion-to-occupancy-driven active distillation extracts rich semantic information from multimodal data and selectively transfers knowledge to image features based on LiDAR-identified regions. Finally, for optimal performance, we introduce SDG-Fusion, which uses fusion alone, and SDG-KL, which integrates both fusion and distillation for faster inference. Our method achieves state-of-the-art (SOTA) performance with real-time processing on the Occ3D-nuScenes dataset and shows comparable performance on the more challenging SurroundOcc-nuScenes dataset, demonstrating its effectiveness and robustness. The code will be released at https://github.com/DzpLab/SDGOCC.

View Paper

Weather-Aware AI Systems versus Route-Optimization AI: A Comprehensive Analysis of AI Applications in Transportation Productivity

Authors: Tatsuru Kikuchi

Abstract: arXiv:2507.17099v1 Announce Type: cross Abstract: While recent research demonstrates that AI route-optimization systems improve taxi driver productivity by 14\%, this study reveals that such findings capture only a fraction of AI's potential in transportation. We examine comprehensive weather-aware AI systems that integrate deep learning meteorological prediction with machine learning positioning optimization, comparing their performance against traditional operations and route-only AI approaches. Using simulation data from 10,000 taxi operations across varied weather conditions, we find that weather-aware AI systems increase driver revenue by 107.3\%, compared to 14\% improvements from route-optimization alone. Weather prediction contributes the largest individual productivity gain, with strong correlations between meteorological conditions and demand ($r=0.575$). Economic analysis reveals annual earnings increases of 13.8 million yen per driver, with rapid payback periods and superior return on investment. These findings suggest that current AI literature significantly underestimates AI's transformative potential by focusing narrowly on routing algorithms, while weather intelligence represents an untapped \$8.9 billion market opportunity. Our results indicate that future AI implementations should adopt comprehensive approaches that address multiple operational challenges simultaneously rather than optimizing isolated functions.

View Paper

Reinforcement Learning Fine-Tunes a Sparse Subnetwork in Large Language Models

Authors: Andrii Balashov

Abstract: arXiv:2507.17107v1 Announce Type: cross Abstract: Reinforcement learning (RL) is a key post-pretraining step for aligning large language models (LLMs) with complex tasks and human preferences. While it is often assumed that RL fine-tuning requires updating most of a model's parameters, we challenge this assumption with a surprising finding: RL fine-tuning consistently modifies only a small subnetwork (typically 5-30% of weights), leaving most parameters unchanged. We call this phenomenon RL-induced parameter update sparsity. It arises naturally, without any sparsity constraints or parameter-efficient tuning, and appears across multiple RL algorithms (e.g., PPO, DPO, SimPO, PRIME) and model families (e.g., OpenAI, Meta, and open-source LLMs). Moreover, the subnetworks updated by RL show substantial overlap across different seeds, datasets, and algorithms-far exceeding chance-suggesting a partially transferable structure in the pretrained model. We show that fine-tuning only this sparse subnetwork recovers full model performance and yields parameters nearly identical to the fully fine-tuned model. Our analysis suggests this sparsity emerges because RL operates near the model's original distribution, requiring only targeted changes. KL penalties, gradient clipping, and on-policy dynamics have limited effect on the sparsity pattern. These findings shed new light on how RL adapts models: not by shifting all weights, but by focusing training on a small, consistently updated subnetwork. This insight enables more efficient RL methods and reframes sparsity through the lens of the lottery ticket hypothesis.

View Paper

BucketServe: Bucket-Based Dynamic Batching for Smart and Efficient LLM Inference Serving

Authors: Wanyi Zheng, Minxian Xu, Shengye Song, Kejiang Ye

Abstract: arXiv:2507.17120v1 Announce Type: cross Abstract: Large language models (LLMs) have become increasingly popular in various areas, traditional business gradually shifting from rule-based systems to LLM-based solutions. However, the inference of LLMs is resource-intensive or latency-sensitive, posing significant challenges for serving systems. Existing LLM serving systems often use static or continuous batching strategies, which can lead to inefficient GPU memory utilization and increased latency, especially under heterogeneous workloads. These methods may also struggle to adapt to dynamic workload fluctuations, resulting in suboptimal throughput and potential service level objective (SLO) violations. In this paper, we introduce BucketServe, a bucket-based dynamic batching framework designed to optimize LLM inference performance. By grouping requests into size-homogeneous buckets based on sequence length, BucketServe minimizes padding overhead and optimizes GPU memory usage through real-time batch size adjustments preventing out-of-memory (OOM) errors. It introduces adaptive bucket splitting/merging and priority-aware scheduling to mitigate resource fragmentation and ensure SLO compliance. Experiment shows that BucketServe significantly outperforms UELLM in throughput, achieving up to 3.58x improvement. It can also handle 1.93x more request load under the SLO attainment of 80% compared with DistServe and demonstrates 1.975x higher system load capacity compared to the UELLM.

View Paper

Enabling Self-Improving Agents to Learn at Test Time With Human-In-The-Loop Guidance

Authors: Yufei He, Ruoyu Li, Alex Chen, Yue Liu, Yulin Chen, Yuan Sui, Cheng Chen, Yi Zhu, Luca Luo, Frank Yang, Bryan Hooi

Abstract: arXiv:2507.17131v1 Announce Type: cross Abstract: Large language model (LLM) agents often struggle in environments where rules and required domain knowledge frequently change, such as regulatory compliance and user risk screening. Current approaches, like offline fine-tuning and standard prompting, are insufficient because they cannot effectively adapt to new knowledge during actual operation. To address this limitation, we propose the Adaptive Reflective Interactive Agent (ARIA), an LLM agent framework designed specifically to continuously learn updated domain knowledge at test time. ARIA assesses its own uncertainty through structured self-dialogue, proactively identifying knowledge gaps and requesting targeted explanations or corrections from human experts. It then systematically updates an internal, timestamped knowledge repository with provided human guidance, detecting and resolving conflicting or outdated knowledge through comparisons and clarification queries. We evaluate ARIA on the realistic customer due diligence name screening task on TikTok Pay, alongside publicly available dynamic knowledge tasks. Results demonstrate significant improvements in adaptability and accuracy compared to baselines using standard offline fine-tuning and existing self-improving agents. ARIA is deployed within TikTok Pay serving over 150 million monthly active users, confirming its practicality and effectiveness for operational use in rapidly evolving environments.

View Paper

Resilient Multi-Agent Negotiation for Medical Supply Chains:Integrating LLMs and Blockchain for Transparent Coordination

Authors: Mariam ALMutairi, Hyungmin Kim

Abstract: arXiv:2507.17134v1 Announce Type: cross Abstract: Global health emergencies, such as the COVID-19 pandemic, have exposed critical weaknesses in traditional medical supply chains, including inefficiencies in resource allocation, lack of transparency, and poor adaptability to dynamic disruptions. This paper presents a novel hybrid framework that integrates blockchain technology with a decentralized, large language model (LLM) powered multi-agent negotiation system to enhance the resilience and accountability of medical supply chains during crises. In this system, autonomous agents-representing manufacturers, distributors, and healthcare institutions-engage in structured, context-aware negotiation and decision-making processes facilitated by LLMs, enabling rapid and ethical allocation of scarce medical resources. The off-chain agent layer supports adaptive reasoning and local decision-making, while the on-chain blockchain layer ensures immutable, transparent, and auditable enforcement of decisions via smart contracts. The framework also incorporates a formal cross-layer communication protocol to bridge decentralized negotiation with institutional enforcement. A simulation environment emulating pandemic scenarios evaluates the system's performance, demonstrating improvements in negotiation efficiency, fairness of allocation, supply chain responsiveness, and auditability. This research contributes an innovative approach that synergizes blockchain trust guarantees with the adaptive intelligence of LLM-driven agents, providing a robust and scalable solution for critical supply chain coordination under uncertainty.

View Paper

SADA: Stability-guided Adaptive Diffusion Acceleration

Authors: Ting Jiang, Yixiao Wang, Hancheng Ye, Zishan Shao, Jingwei Sun, Jingyang Zhang, Zekai Chen, Jianyi Zhang, Yiran Chen, Hai Li

Abstract: arXiv:2507.17135v1 Announce Type: cross Abstract: Diffusion models have achieved remarkable success in generative tasks but suffer from high computational costs due to their iterative sampling process and quadratic attention costs. Existing training-free acceleration strategies that reduce per-step computation cost, while effectively reducing sampling time, demonstrate low faithfulness compared to the original baseline. We hypothesize that this fidelity

gap arises because (a) different prompts correspond to varying denoising trajectory, and (b) such methods do not consider the underlying ODE formulation and its numerical solution. In this paper, we propose Stability-guided Adaptive Diffusion Acceleration (SADA), a novel paradigm that unifies step-wise and token-wise sparsity decisions via a single stability criterion to accelerate sampling of ODE-based generative models (Diffusion and Flow-matching). For (a), SADA adaptively allocates sparsity based on the sampling trajectory. For (b), SADA introduces principled approximation schemes that leverage the precise gradient information from the numerical ODE solver. Comprehensive evaluations on SD-2, SDXL, and Flux using both EDM and DPM++ solvers reveal consistent $\ge 1.8\times$ speedups with minimal fidelity degradation (LPIPS ≤ 0.10 and FID ≤ 4.5) compared to unmodified baselines, significantly outperforming prior methods. Moreover, SADA adapts seamlessly to other pipelines and modalities: It accelerates ControlNet without any modifications and speeds up MusicLDM by $1.8\times$ with ~ 0.01 spectrogram LPIPS.

View Paper

Towards Human-level Intelligence via Human-like Whole-Body Manipulation

Authors: Guang Gao, Jianan Wang, Jinbo Zuo, Junnan Jiang, Jingfan Zhang, Xianwen Zeng, Yuejiang Zhu, Lianyang Ma, Ke Chen, Minhua Sheng, Ruirui Zhang, Zhaohui An

Abstract: arXiv:2507.17141v1 Announce Type: cross Abstract: Building general-purpose intelligent robots has long been a fundamental goal of robotics. A promising approach is to mirror the evolutionary trajectory of humans: learning through continuous interaction with the environment, with early progress driven by the imitation of human behaviors. Achieving this goal presents three core challenges: (1) designing safe robotic hardware with human-level physical capabilities; (2) developing an intuitive and scalable whole-body teleoperation interface for data collection; and (3) creating algorithms capable of learning whole-body visuomotor policies from human demonstrations. To address these challenges in a unified framework, we propose Astribot Suite, a robot learning suite for whole-body manipulation aimed at general daily tasks across diverse environments. We demonstrate the effectiveness of our system on a wide range of activities that require whole-body coordination, extensive reachability, human-level dexterity, and agility. Our results show that Astribot's cohesive integration of embodiment, teleoperation interface, and learning pipeline marks a significant step towards real-world, general-purpose whole-body robotic manipulation, laying the groundwork for the next generation of intelligent robots.

View Paper

ScSAM: Debiasing Morphology and Distributional Variability in Subcellular Semantic Segmentation

Authors: Bo Fang, Jianan Fan, Dongnan Liu, Hang Chang, Gerald J. Shami, Filip Braet, Weidong Cai

Abstract: arXiv:2507.17149v1 Announce Type: cross Abstract: The significant morphological and distributional variability among subcellular components poses a long-standing challenge for learning-based organelle segmentation models, significantly increasing the risk of biased feature learning. Existing methods often rely on single mapping relationships, overlooking feature diversity and thereby inducing biased training. Although the Segment Anything Model (SAM) provides rich feature representations, its application to subcellular scenarios is hindered by two key challenges: (1) The variability in subcellular morphology and distribution creates gaps in the label space, leading the model to learn spurious or biased features. (2) SAM focuses on global contextual understanding and often ignores fine-grained spatial details, making it challenging to capture subtle structural alterations and cope with skewed data distributions. To address these challenges, we introduce ScSAM, a method that enhances feature robustness by fusing pre-trained SAM with Masked Autoencoder (MAE)-guided cellular prior knowledge to alleviate training bias from data imbalance. Specifically, we design a feature alignment and fusion module to align pre-trained embeddings to the same feature space and efficiently combine different representations. Moreover, we present a cosine similarity matrix-based class prompt encoder to activate class-specific features to recognize subcellular categories. Extensive experiments on diverse subcellular image datasets demonstrate that ScSAM outperforms state-of-the-art methods.

View Paper

JAM: Keypoint-Guided Joint Prediction after Classification-Aware Marginal Proposal for Multi-Agent Interaction

Authors: Fangze Lin, Ying He, Fei Yu, Hong Zhang

Abstract: arXiv:2507.17152v1 Announce Type: cross Abstract: Predicting the future motion of road participants is a critical task in autonomous driving. In this work, we address the challenge of low-quality generation of low-probability modes in multi-agent joint prediction. To tackle this issue, we propose a two-stage multi-agent interactive prediction framework named \textit{keypoint-guided joint prediction after classification-aware marginal proposal} (JAM). The first stage is modeled as a marginal prediction process, which classifies queries by trajectory type to encourage the model to learn all categories of trajectories, providing comprehensive mode information for the joint prediction module. The second stage is modeled as a joint prediction process, which takes the scene context and the marginal proposals from the first stage as inputs to learn the final joint distribution. We explicitly introduce key waypoints to guide the joint prediction module in better capturing and leveraging the critical information from the initial predicted trajectories. We conduct extensive experiments on the real-world Waymo Open Motion Dataset interactive prediction benchmark. The results show that our approach achieves competitive performance. In particular, in the framework comparison experiments, the proposed JAM outperforms other prediction frameworks and achieves state-of-the-art performance in interactive trajectory prediction. The code is available at https://github.com/LinFunster/JAM to facilitate future research.

View Paper

Tabular Diffusion based Actionable Counterfactual Explanations for Network Intrusion Detection

Authors: Vinura Galwaduge, Jagath Samarabandu

Abstract: arXiv:2507.17161v1 Announce Type: cross Abstract: Modern network intrusion detection systems (NIDS) frequently utilize the predictive power of complex deep learning models. However, the "black-box" nature of such deep learning methods adds a layer of opaqueness that hinders the proper understanding of detection decisions, trust in the decisions and prevent timely countermeasures against such attacks. Explainable AI (XAI) methods provide a solution to this problem by providing insights into the causes of the predictions. The majority of the existing XAI methods provide explanations which are not convenient to convert into actionable countermeasures. In this work, we propose a novel diffusion-based counterfactual explanation framework that can provide actionable explanations for network intrusion attacks. We evaluated our proposed algorithm against several other publicly available counterfactual explanation algorithms on 3 modern network intrusion datasets. To the best of our knowledge, this work also presents the first comparative analysis of existing counterfactual explanation algorithms within the context of network intrusion detection systems. Our proposed method provide minimal, diverse counterfactual explanations out of the tested counterfactual explanation algorithms in a more efficient manner by reducing the time to generate explanations. We also demonstrate how counterfactual explanations can provide actionable explanations by summarizing them to create a set of global rules. These rules are actionable not only at instance level but also at the global level for intrusion attacks. These global counterfactual rules show the ability to effectively filter out incoming attack queries which is crucial for efficient intrusion detection and defense mechanisms.

View Paper

SKA-Bench: A Fine-Grained Benchmark for Evaluating Structured Knowledge Understanding of LLMs

Authors: Zhiqiang Liu, Enpei Niu, Yin Hua, Mengshu Sun, Lei Liang, Huajun Chen, Wen Zhang

Abstract: arXiv:2507.17178v1 Announce Type: cross Abstract: Although large language models (LLMs) have made significant progress in understanding Structured Knowledge (SK) like KG and Table, existing evaluations for SK understanding are non-rigorous (i.e., lacking evaluations of specific capabilities) and focus on a single type of SK. Therefore, we aim to propose a more comprehensive and rigorous structured knowledge understanding benchmark to diagnose the shortcomings of LLMs. In this paper, we introduce SKA-Bench, a Structured Knowledge Augmented QA Benchmark that encompasses four widely used structured knowledge forms: KG, Table, KG+Text, and Table+Text. We utilize a three-stage pipeline to construct SKA-Bench instances, which includes a question, an answer, positive knowledge units, and noisy knowledge units. To evaluate the SK understanding capabilities of LLMs in a fine-grained manner, we expand the instances into four fundamental ability testbeds: Noise Robustness, Order Insensitivity, Information Integration, and Negative Rejection. Empirical evaluations on 8 representative LLMs, including the advanced DeepSeek-R1, indicate that existing LLMs still face significant challenges in understanding structured knowledge, and their performance is influenced by factors such as the amount of noise, the order of knowledge units, and hallucination phenomenon. Our dataset and code are available at https://github.com/Lza12a/SKA-Bench.

View Paper

Regret Minimization in Population Network Games: Vanishing Heterogeneity and Convergence to Equilibria

Authors: Die Hu, Shuyue Hu, Chunjiang Mu, Shiqi Fan, Chen Chu, Jinzhuo Liu, Zhen Wang

Abstract: arXiv:2507.17183v1 Announce Type: cross Abstract: Understanding and predicting the behavior of large-scale multi-agents in games remains a fundamental challenge in multi-agent systems. This paper examines the role of heterogeneity in equilibrium formation by analyzing how smooth regret-matching drives a large number of heterogeneous agents with diverse initial policies toward unified behavior. By modeling the system state as a probability distribution of regrets and analyzing its evolution through the continuity equation, we uncover a key phenomenon in diverse multi-agent settings: the variance of the regret distribution diminishes over time, leading to the disappearance of heterogeneity and the emergence of consensus among agents. This universal result enables us to prove convergence to quantal response equilibria in both competitive and cooperative multi-agent settings. Our work advances the theoretical understanding of multi-agent learning and offers a novel perspective on equilibrium selection in diverse game-theoretic scenarios.

View Paper

Asymmetric Lesion Detection with Geometric Patterns and CNN-SVM Classification

Authors: M. A. Rasel, Sameem Abdul Kareem, Zhenli Kwan, Nik Aimee Azizah Faheem, Winn Hui Han, Rebecca Kai Jan Choong, Shin Shen Yong, Unaizah Obaidellah

Abstract: arXiv:2507.17185v1 Announce Type: cross Abstract: In dermoscopic images, which allow visualization of surface skin structures not visible to the naked eye, lesion shape offers vital insights into skin diseases. In clinically practiced methods, asymmetric lesion shape is one of the criteria for diagnosing melanoma. Initially, we labeled data for a non-annotated dataset with symmetrical information based on clinical assessments. Subsequently, we propose a supporting technique, a supervised learning image processing algorithm, to analyze the geometrical pattern of lesion shape, aiding non-experts in understanding the criteria of an asymmetric lesion. We then utilize a pre-trained convolutional neural network (CNN) to extract shape, color, and texture features from dermoscopic images for training a multiclass support vector machine (SVM) classifier, outperforming state-of-the-art methods from the literature. In the geometry-based experiment, we achieved a 99.00% detection rate for dermatological asymmetric lesions. In the CNN-based experiment, the best performance is found with 94% Kappa Score, 95% Macro F1-score, and 97% Weighted F1-score for classifying lesion shapes (Asymmetric, Half-Symmetric, and Symmetric).

View Paper

LLM Meets the Sky: Heuristic Multi-Agent Reinforcement Learning for Secure Heterogeneous UAV Networks

Authors: Lijie Zheng, Ji He, Shih Yu Chang, Yulong Shen, Dusit Niyato

Abstract: arXiv:2507.17188v1 Announce Type: cross Abstract: This work tackles the physical layer security (PLS) problem of maximizing the secrecy rate in heterogeneous UAV networks (HetUAVNs) under propulsion energy constraints. Unlike prior studies that assume uniform UAV capabilities or overlook energy-security trade-offs, we consider a realistic scenario where UAVs with diverse payloads and computation resources collaborate to serve ground terminals in the presence of eavesdroppers. To manage the complex coupling between UAV motion and communication, we propose a hierarchical optimization framework. The inner layer uses a semidefinite relaxation (SDR)-based S2DC algorithm combining penalty functions and difference-of-convex (d.c.) programming to solve the secrecy precoding problem with fixed UAV positions. The outer layer introduces a Large Language Model (LLM)-guided heuristic multi-agent reinforcement learning approach (LLM-HeMARL) for trajectory optimization. LLM-HeMARL efficiently incorporates expert heuristics policy generated by the LLM, enabling UAVs to learn energy-aware, security-driven trajectories without the inference overhead of real-time LLM calls. The simulation results show that our method outperforms existing baselines in secrecy rate and energy efficiency, with consistent robustness across varying UAV swarm sizes and random seeds.

View Paper

Dispatch-Aware Deep Neural Network for Optimal Transmission Switching: Toward Real-Time and Feasibility Guaranteed Operation

Authors: Minsoo Kim, Jip Kim

Abstract: arXiv:2507.17194v1 Announce Type: cross Abstract: Optimal transmission switching (OTS) improves optimal power flow (OPF) by selectively opening transmission lines, but its mixed-integer formulation increases computational complexity, especially on large grids. To deal with this, we propose a dispatch-aware deep neural network (DA-DNN) that accelerates DC-OTS without relying on pre-solved labels. DA-DNN predicts line states and passes them through a differentiable DC-OPF layer, using the resulting generation cost as the loss function so that all physical network constraints are enforced throughout training and inference. In addition, we adopt a customized weight-bias initialization that keeps every forward pass feasible from the first iteration, which allows stable learning on large grids. Once trained, the proposed DA-DNN produces a provably feasible topology and dispatch pair in the same time as solving the DCOPF, whereas conventional mixed-integer solvers become intractable. As a result, the proposed method successfully captures the economic advantages of OTS while maintaining scalability.

View Paper

DesignLab: Designing Slides Through Iterative Detection and Correction

Authors: Jooyeol Yun, Heng Wang, Yotaro Shimose, Jaegul Choo, Shingo Takamatsu

Abstract: arXiv:2507.17202v1 Announce Type: cross Abstract: Designing high-quality presentation slides can be challenging for non-experts due to the complexity involved in navigating various design choices. Numerous automated tools can suggest layouts and color schemes, yet often lack the ability to refine their own output, which is a key aspect in real-world workflows. We propose DesignLab, which separates the design process into two roles, the design reviewer, who identifies design-related issues, and the design contributor who corrects them. This decomposition enables an iterative loop where the reviewer continuously detects issues and the contributor corrects them, allowing a draft to be further polished with each iteration, reaching qualities that were unattainable. We fine-tune large language models for these roles and simulate intermediate drafts by introducing controlled perturbations, enabling the design reviewer learn design errors and the contributor learn how to fix them. Our experiments show that DesignLab outperforms existing design-generation methods, including a commercial tool, by embracing the iterative nature of designing which can result in polished, professional slides.

View Paper

The Pluralistic Moral Gap: Understanding Judgment and Value Differences between Humans and Large Language Models

Authors: Giuseppe Russo, Debora Nozza, Paul R\"ottger, Dirk Hovy

Abstract: arXiv:2507.17216v1 Announce Type: cross Abstract: People increasingly rely on Large Language Models (LLMs) for moral advice, which may influence humans' decisions. Yet, little is known about how closely LLMs align

with human moral judgments. To address this, we introduce the Moral Dilemma Dataset, a benchmark of 1,618 real-world moral dilemmas paired with a distribution of human moral judgments consisting of a binary evaluation and a free-text rationale. We treat this problem as a pluralistic distributional alignment task, comparing the distributions of LLM and human judgments across dilemmas. We find that models reproduce human judgments only under high consensus; alignment deteriorates sharply when human disagreement increases. In parallel, using a 60-value taxonomy built from 3,783 value expressions extracted from rationales, we show that LLMs rely on a narrower set of moral values than humans. These findings reveal a pluralistic moral gap: a mismatch in both the distribution and diversity of values expressed. To close this gap, we introduce Dynamic Moral Profiling (DMP), a Dirichlet-based sampling method that conditions model outputs on human-derived value profiles. DMP improves alignment by 64.3% and enhances value diversity, offering a step toward more pluralistic and human-aligned moral guidance from LLMs.

View Paper

HuiduRep: A Robust Self-Supervised Framework for Learning Neural Representations from Extracellular Spikes

Authors: Feng Cao, Zishuo Feng

Abstract: arXiv:2507.17224v1 Announce Type: cross Abstract: Extracellular recordings are brief voltage fluctuations recorded near neurons, widely used in neuroscience as the basis for decoding brain activity at single-neuron resolution. Spike sorting, which assigns each spike to its source neuron, is a critical step in brain sensing pipelines. However, it remains challenging under low signal-to-noise ratio (SNR), electrode drift, and cross-session variability. In this paper, we propose HuiduRep, a robust self-supervised representation learning framework that extracts discriminative and generalizable features from extracellular spike waveforms. By combining contrastive learning with a denoising autoencoder, HuiduRep learns latent representations that are robust to noise and drift. Built on HuiduRep, we develop a spike sorting pipeline that clusters spike representations without supervision. Experiments on hybrid and real-world datasets demonstrate that HuiduRep achieves strong robustness and the pipeline matches or outperforms state-of-the-art tools such as KiloSort4 and MountainSort5. These findings demonstrate the potential of self-supervised spike representation learning as a foundational tool for robust and generalizable processing of extracellular recordings.

View Paper

P3SL: Personalized Privacy-Preserving Split Learning on Heterogeneous Edge Devices

Authors: Wei Fan, JinYi Yoon, Xiaochang Li, Huajie Shao, Bo Ji

Abstract: arXiv:2507.17228v1 Announce Type: cross Abstract: Split Learning (SL) is an emerging privacy-preserving machine learning technique that enables resource constrained edge devices to participate in model training by partitioning a model into client-side and server-side sub-models. While SL reduces computational overhead on edge devices, it encounters significant challenges in heterogeneous environments where devices vary in computing resources, communication capabilities, environmental conditions, and privacy requirements. Although recent studies have explored heterogeneous SL frameworks that optimize split points for devices with varying resource constraints, they often neglect personalized privacy requirements and local model customization under varying environmental conditions. To address these limitations, we propose P3SL, a Personalized Privacy-Preserving Split Learning framework designed for heterogeneous, resource-constrained edge device systems. The key contributions of this work are twofold. First, we design a personalized sequential split learning pipeline that allows each client to achieve customized privacy protection and maintain personalized local models tailored to their computational resources, environmental conditions, and privacy needs. Second, we adopt a bi-level optimization technique that empowers clients to determine their own optimal personalized split points without sharing private sensitive information (i.e., computational resources, environmental conditions, privacy requirements) with the server. This approach balances energy consumption and privacy leakage risks while maintaining high model accuracy. We implement and evaluate P3SL on a testbed consisting of 7 devices including 4 Jetson Nano P3450 devices, 2 Raspberry Pis, and 1 laptop, using diverse model architectures and datasets under varying environmental conditions.

View Paper

A Highly Clean Recipe Dataset with Ingredient States Annotation for State Probing Task

Authors: Mashiro Toyooka, Kiyoharu Aizawa, Yoko Yamakata

Abstract: arXiv:2507.17232v1 Announce Type: cross Abstract: Large Language Models (LLMs) are trained on a vast amount of procedural texts, but they do not directly observe real-world phenomena. In the context of cooking recipes, this poses a challenge, as intermediate states of ingredients are often omitted, making it difficult for models to track ingredient states and understand recipes accurately. In this paper, we apply state probing, a method for evaluating a language model's understanding of the world, to the domain of cooking. We propose a new task and dataset for evaluating how well LLMs can recognize intermediate ingredient states during cooking procedures. We first construct a new Japanese recipe dataset with clear and accurate annotations of ingredient state changes, collected from well-structured and controlled recipe texts. Using this dataset, we design three novel tasks to evaluate whether LLMs can track ingredient state transitions and identify ingredients present at intermediate steps. Our experiments with widely used LLMs, such as Llama3.1-70B and Qwen2.5-72B, show that learning ingredient state knowledge improves their understanding of cooking processes, achieving performance comparable to commercial LLMs.

View Paper

Eco-Friendly AI: Unleashing Data Power for Green Federated Learning

Authors: Mattia Sabella, Monica Vitali

Abstract: arXiv:2507.17241v1 Announce Type: cross Abstract: The widespread adoption of Artificial Intelligence (AI) and Machine Learning (ML) comes with a significant environmental impact, particularly in terms of energy consumption and carbon emissions. This pressing issue highlights the need for innovative solutions to mitigate AI's ecological footprint. One of the key factors influencing the energy consumption of ML model training is the size of the training dataset. ML models are often trained on vast amounts of data continuously generated by sensors and devices distributed across multiple locations. To reduce data transmission costs and enhance privacy, Federated Learning (FL) enables model training without the need to move or share raw data. While FL offers these advantages, it also introduces challenges due to the heterogeneity of data sources (related to volume and quality), computational node capabilities, and environmental impact. This paper contributes to the advancement of Green AI by proposing a data-centric approach to Green Federated Learning. Specifically, we focus on reducing FL's environmental impact by minimizing the volume of training data. Our methodology involves the analysis of the characteristics of federated datasets, the selecting of an optimal subset of data based on quality metrics, and the choice of the federated nodes with the lowest environmental impact. We develop a comprehensive methodology that examines the influence of data-centric factors, such as data quality and volume, on FL training performance and carbon emissions. Building on these insights, we introduce an interactive recommendation system that optimizes FL configurations through data reduction, minimizing environmental impact during training. Applying this methodology to time series classification has demonstrated promising results in reducing the environmental impact of FL tasks.

View Paper

DistrAttention: An Efficient and Flexible Self-Attention Mechanism on Modern GPUs

Authors: Haolin Jin, Mengbai Xiao, Yuan Yuan, Xiao Zhang, Dongxiao Yu, Guanghui Zhang, Haoliang Wang

Abstract: arXiv:2507.17245v1 Announce Type: cross Abstract: The Transformer architecture has revolutionized deep learning, delivering the state-of-the-art performance in areas such as natural language processing, computer vision, and time series prediction. However, its core component, self-attention, has the quadratic time complexity relative to input sequence length, which hinders the scalability of Transformers. The exsiting approaches on optimizing self-attention either discard full-contextual information or lack of flexibility. In this work, we design DistrAttention, an effcient and flexible self-attention mechanism with the full context. DistrAttention achieves this by grouping data on the embedding dimensionality, usually referred to as d. We realize DistrAttention with a lightweight sampling and fusion method that exploits locality-sensitive hashing to group similar data. A block-wise grouping framework is further designed to limit the errors introduced by locality sensitive hashing. By optimizing the selection of block sizes, DistrAttention could be easily integrated with FlashAttention-2, gaining high-performance on modern GPUs. We evaluate DistrAttention with extensive experiments. The results show that our method is 37% faster than FlashAttention-2 on calculating self-attention. In ViT inference, DistrAttention is the fastest and the most accurate among approximate self-attention mechanisms. In Llama3-1B, DistrAttention still achieves the lowest inference time with only 1% accuray loss.

View Paper

Reality Proxy: Fluid Interactions with Real-World Objects in MR via Abstract Representations

Authors: Xiaoan Liu, Difan Jia, Xianhao Carton Liu, Mar Gonzalez-Franco, Chen Zhu-Tian

Abstract: arXiv:2507.17248v1 Announce Type: cross Abstract: Interacting with real-world objects in Mixed Reality (MR) often proves difficult when they are crowded, distant, or partially occluded, hindering straightforward selection and manipulation. We observe that these difficulties stem from performing interaction directly on physical objects, where input is tightly coupled to their physical constraints. Our key insight is to decouple interaction from these constraints by introducing proxies-abstract representations of real-world objects. We embody this concept in Reality Proxy, a system that seamlessly shifts interaction targets from physical objects to their proxies during selection. Beyond facilitating basic selection, Reality Proxy uses AI to enrich proxies with semantic attributes and hierarchical spatial relationships of their corresponding physical objects, enabling novel and previously cumbersome interactions in MR - such as skimming, attribute-based filtering, navigating nested groups, and complex multi object selections - all without requiring new gestures or menu systems. We demonstrate Reality Proxy's versatility across diverse scenarios, including office information retrieval, large-scale spatial navigation, and multi-drone control. An expert evaluation suggests the system's utility and usability, suggesting that proxy-based abstractions offer a powerful and generalizable interaction paradigm for future MR systems.

View Paper

Understanding Prompt Programming Tasks and Questions

Authors: Jenny T. Liang, Chenyang Yang, Agnia Sergeyuk, Travis D. Breaux, Brad A. Myers

Abstract: arXiv:2507.17264v1 Announce Type: cross Abstract: Prompting foundation models (FMs) like large language models (LLMs) have enabled new AI-powered software features (e.g., text summarization) that previously were only possible by fine-tuning FMs. Now, developers are embedding prompts in software, known as prompt programs. The process of prompt programming requires the developer to make many changes to their prompt. Yet, the questions developers ask to update their prompt is unknown, despite the answers to these questions affecting how developers plan their changes. With the growing number of research and commercial prompt programming tools, it is unclear whether prompt programmers' needs are being adequately addressed. We address these challenges by developing a taxonomy of 25 tasks prompt programmers do and 51 questions they ask, measuring the importance of each task and question. We interview 16 prompt programmers, observe 8 developers make prompt changes, and survey 50 developers. We then compare the taxonomy with 48 research and commercial tools. We find that prompt programming is not well-supported: all tasks are done manually, and 16 of the 51 questions -- including a majority of the most important ones -- remain unanswered. Based on this, we outline important opportunities for prompt programming tools.

View Paper

Leveraging Knowledge Graphs and LLM Reasoning to Identify Operational Bottlenecks for Warehouse Planning Assistance

Authors: Rishi Parekh, Saisubramaniam Gopalakrishnan, Zishan Ahmad, Anirudh Deodhar

Abstract: arXiv:2507.17273v1 Announce Type: cross Abstract: Analyzing large, complex output datasets from Discrete Event Simulations (DES) of warehouse operations to identify bottlenecks and inefficiencies is a critical yet challenging task, often demanding significant manual effort or specialized analytical tools. Our framework integrates Knowledge Graphs (KGs) and Large Language Model (LLM)-based agents to analyze complex Discrete Event Simulation (DES) output data from warehouse operations. It transforms raw DES data into a semantically rich KG, capturing relationships between simulation events and entities. An LLM-based agent uses iterative reasoning, generating interdependent sub-questions. For each sub-question, it creates Cypher queries for KG interaction, extracts information, and self-reflects to correct errors. This adaptive, iterative, and self-correcting process identifies operational issues mimicking human analysis. Our DES approach for warehouse bottleneck identification, tested with equipment breakdowns and process irregularities, outperforms baseline methods. For operational questions, it achieves near-perfect pass rates in pinpointing inefficiencies. For complex investigative questions, we demonstrate its superior diagnostic ability to uncover subtle, interconnected issues. This work bridges simulation modeling and AI (KG+LLM), offering a more intuitive method for actionable insights, reducing time-to-insight, and enabling automated warehouse inefficiency evaluation and diagnosis.

View Paper

Integrating Belief Domains into Probabilistic Logic Programs

Authors: Damiano Azzolini, Fabrizio Riguzzi, Theresa Swift

Abstract: arXiv:2507.17291v1 Announce Type: cross Abstract: Probabilistic Logic Programming (PLP) under the Distribution Semantics is a leading approach to practical reasoning under uncertainty. An advantage of the Distribution Semantics is its suitability for implementation as a Prolog or Python library, available through two well-maintained implementations, namely ProbLog and cplint/PITA. However, current formulations of the Distribution Semantics use point-probabilities, making it difficult to express epistemic uncertainty, such as arises from, for example, hierarchical classifications from computer vision models. Belief functions generalize probability measures as non-additive capacities, and address epistemic uncertainty via interval probabilities. This paper introduces interval-based Capacity Logic Programs based on an extension of the Distribution Semantics to include belief functions, and describes properties of the new framework that make it amenable to practical applications.

View Paper

On Temporal Guidance and Iterative Refinement in Audio Source Separation

Authors: Tobias Morocutti, Jonathan Greif, Paul Primus, Florian Schmid, Gerhard Widmer

Abstract: arXiv:2507.17297v1 Announce Type: cross Abstract: Spatial semantic segmentation of sound scenes (S5) involves the accurate identification of active sound classes and the precise separation of their sources from complex acoustic mixtures. Conventional systems rely on a two-stage pipeline - audio tagging followed by label-conditioned source separation - but are often constrained by the absence of fine-grained temporal information critical for effective separation. In this work, we address this limitation by introducing a novel approach for S5 that enhances the synergy between the event detection and source separation stages. Our key contributions are threefold. First, we fine-tune a pre-trained Transformer to detect active sound classes. Second, we utilize a separate instance of this fine-tuned Transformer to perform sound event detection (SED), providing the separation module with detailed, time-varying guidance. Third, we implement an iterative refinement mechanism that progressively enhances separation quality by recursively reusing the separator's output from previous iterations. These advancements lead to significant improvements in both audio tagging and source separation performance, as demonstrated by our system's second-place finish in Task 4 of the DCASE Challenge 2025. Our implementation and model checkpoints are available in our GitHub repository: https://github.com/theMoro/dcase25task4 .

View Paper

A Versatile Pathology Co-pilot via Reasoning Enhanced Multimodal Large Language Model

Authors: Zhe Xu, Ziyi Liu, Junlin Hou, Jiabo Ma, Cheng Jin, Yihui Wang, Zhixuan Chen, Zhengyu Zhang, Zhengrui Guo, Fengtao Zhou, Yingxue Xu, Xi Wang, Ronald Cheong Kin Chan, Li Liang, Hao Chen

Abstract: arXiv:2507.17303v1 Announce Type: cross Abstract: Multimodal large language models (MLLMs) have emerged as powerful tools for computational pathology, offering unprecedented opportunities to integrate pathological images with language context for comprehensive diagnostic analysis. These models hold particular promise for automating complex tasks that traditionally require expert interpretation of pathologists. However, current MLLM approaches in pathology demonstrate significantly constrained reasoning capabilities, primarily due to their reliance on expensive chain-of-thought annotations. Additionally, existing methods remain limited to simplex application of visual question answering (VQA) at region-of-interest (ROI) level, failing to address the full spectrum of diagnostic needs such as ROI classification, detection, segmentation, whole-slide-image (WSI) classification and VQA in clinical practice. In this study, we present SmartPath-R1, a versatile MLLM capable of simultaneously addressing both ROI-level and WSI-level tasks while demonstrating robust pathological reasoning capability. Our framework combines scale-dependent supervised fine-tuning and task-aware reinforcement fine-tuning, which circumvents the requirement for chain-of-thought supervision by leveraging the intrinsic knowledge within MLLM. Furthermore, SmartPath-R1 integrates multiscale and multitask analysis through a mixture-of-experts mechanism, enabling dynamic processing for diverse tasks. We curate a large-scale dataset comprising 2.3M ROI samples and 188K WSI samples for training and evaluation. Extensive experiments across 72 tasks validate the effectiveness and superiority of the proposed approach. This work represents a significant step toward developing versatile, reasoning-enhanced AI systems for precision pathology.

View Paper

Confounded Causal Imitation Learning with Instrumental Variables

Authors: Yan Zeng, Shenglan Nie, Feng Xie, Libo Huang, Peng Wu, Zhi Geng

Abstract: arXiv:2507.17309v1 Announce Type: cross Abstract: Imitation learning from demonstrations usually suffers from the confounding effects of unmeasured variables (i.e., unmeasured confounders) on the states and actions. If ignoring them, a biased estimation of the policy would be entailed. To break up this confounding gap, in this paper, we take the best of the strong power of instrumental variables (IV) and propose a Confounded Causal Imitation Learning (C2L) model. This model accommodates confounders that influence actions across multiple timesteps, rather than being restricted to immediate temporal dependencies. We develop a two-stage imitation learning framework for valid IV identification and policy optimization. In particular, in the first stage, we construct a testing criterion based on the defined pseudo-variable, with which we achieve identifying a valid IV for the C2L models. Such a criterion entails the

sufficient and necessary identifiability conditions for IV validity. In the second stage, with the identified IV, we propose two candidate policy learning approaches: one is based on a simulator, while the other is offline. Extensive experiments verified the effectiveness of identifying the valid IV as well as learning the policy.

View Paper

EarthLink: Interpreting Climate Signals with Self-Evolving AI Agents

Authors: Zijie Guo, Jiong Wang, Xiaoyu Yue, Wangxu Wei, Zhe Jiang, Wanghan Xu, Ben Fei, Wenlong Zhang, Xinyu Gu, Lijing Cheng, Jing-Jia Luo, Chao Li, Yaqiang Wang, Tao Chen, Wanli Ouyang, Fenghua Ling, Lei Bai

Abstract: arXiv:2507.17311v1 Announce Type: cross Abstract: Modern Earth science is at an inflection point. The vast, fragmented, and complex nature of Earth system data, coupled with increasingly sophisticated analytical demands, creates a significant bottleneck for rapid scientific discovery. Here we introduce EarthLink, the first AI agent designed as an interactive copilot for Earth scientists. It automates the end-to-end research workflow, from planning and code generation to multi-scenario analysis. Unlike static diagnostic tools, EarthLink can learn from user interaction, continuously refining its capabilities through a dynamic feedback loop. We validated its performance on a number of core scientific tasks of climate change, ranging from model-observation comparisons to the diagnosis of complex phenomena. In a multi-expert evaluation, EarthLink produced scientifically sound analyses and demonstrated an analytical competency that was rated as comparable to specific aspects of a human junior researcher's workflow. Additionally, its transparent, auditable workflows and natural language interface empower scientists to shift from laborious manual execution to strategic oversight and hypothesis generation. EarthLink marks a pivotal step towards an efficient, trustworthy, and collaborative paradigm for Earth system research in an era of accelerating global change.

View Paper

Temporal Point-Supervised Signal Reconstruction: A Human-Annotation-Free Framework for Weak Moving Target Detection

Authors: Weihua Gao, Chunxu Ren, Wenlong Niu, Xiaodong Peng

Abstract: arXiv:2507.17334v1 Announce Type: cross Abstract: In low-altitude surveillance and early warning systems, detecting weak moving targets remains a significant challenge due to low signal energy, small spatial extent, and complex background clutter. Existing methods struggle with extracting robust features and suffer from the lack of reliable annotations. To address these limitations, we propose a novel Temporal Point-Supervised (TPS) framework that enables high-performance detection of weak targets without any manual annotations.Instead of conventional frame-based detection, our framework reformulates the task as a pixel-wise temporal signal modeling problem, where weak targets manifest as short-duration pulse-like responses. A Temporal Signal Reconstruction Network (TSRNet) is developed under the TPS paradigm to reconstruct these transient signals.TSRNet adopts an encoder-decoder architecture and integrates a Dynamic Multi-Scale Attention (DMSAttention) module to enhance its sensitivity to diverse temporal patterns. Additionally, a graph-based trajectory mining strategy is employed to suppress false alarms and ensure temporal consistency.Extensive experiments on a purpose-built low-SNR dataset demonstrate that our framework outperforms state-of-the-art methods while requiring no human annotations. It achieves strong detection performance and operates at over 1000 FPS, underscoring its potential for real-time deployment in practical scenarios.

View Paper

Swin-TUNA : A Novel PEFT Approach for Accurate Food Image Segmentation

Authors: Haotian Chen, Zhiyong Xiao

Abstract: arXiv:2507.17347v1 Announce Type: cross Abstract: In the field of food image processing, efficient semantic segmentation techniques are crucial for industrial applications. However, existing large-scale Transformer-based models (such as FoodSAM) face challenges in meeting practical deploymentrequirements due to their massive parameter counts and high computational resource demands. This paper introduces TUNable Adapter module (Swin-TUNA), a Parameter Efficient Fine-Tuning (PEFT) method that integrates multiscale trainable adapters into the Swin Transformer architecture, achieving high-performance food image segmentation by updating only 4% of the parameters. The core innovation of Swin-TUNA lies in its hierarchical feature adaptation mechanism: it designs separable convolutions in depth and dimensional mappings of varying scales to address the differences in features between shallow and deep networks, combined with a dynamic balancing strategy for tasks-agnostic and task-specific features. Experiments demonstrate that this method achieves mIoU of 50.56% and 74.94% on the FoodSeg103 and UECFoodPix Complete datasets, respectively, surpassing the fully parameterized FoodSAM model while reducing the parameter count by 98.7% (to only 8.13M).

Furthermore, Swin-TUNA exhibits faster convergence and stronger generalization capabilities in low-data scenarios, providing an efficient solution for assembling lightweight food image.

View Paper

DynaSearcher: Dynamic Knowledge Graph Augmented Search Agent via Multi-Reward Reinforcement Learning

Authors: Chuzhan Hao, Wenfeng Feng, Yuewei Zhang, Hao Wang

Abstract: arXiv:2507.17365v1 Announce Type: cross Abstract: Multi-step agentic retrieval systems based on large language models (LLMs) have demonstrated remarkable performance in complex information search tasks. However, these systems still face significant challenges in practical applications, particularly in generating factually inconsistent intermediate queries and inefficient search trajectories, which can lead to reasoning deviations or redundant computations. To address these issues, we propose DynaSearcher, an innovative search agent enhanced by dynamic knowledge graphs and multi-reward reinforcement learning (RL). Specifically, our system leverages knowledge graphs as external structured knowledge to guide the search process by explicitly modeling entity relationships, thereby ensuring factual consistency in intermediate queries and mitigating biases from irrelevant information. Furthermore, we employ a multi-reward RL framework for fine-grained control over training objectives such as retrieval accuracy, efficiency, and response quality. This framework promotes the generation of high-quality intermediate queries and comprehensive final answers, while discouraging unnecessary exploration and minimizing information omissions or redundancy. Experimental results demonstrate that our approach achieves state-of-the-art answer accuracy on six multi-hop question answering datasets, matching frontier LLMs while using only small-scale models and limited computational resources. Furthermore, our approach demonstrates strong generalization and robustness across diverse retrieval environments and larger-scale models, highlighting its broad applicability.

View Paper

SFUOD: Source-Free Unknown Object Detection

Authors: Keon-Hee Park, Seun-An Choe, Gyeong-Moon Park

Abstract: arXiv:2507.17373v1 Announce Type: cross Abstract: Source-free object detection adapts a detector pre-trained on a source domain to an unlabeled target domain without requiring access to labeled source data. While this setting is practical as it eliminates the need for the source dataset during domain adaptation, it operates under the restrictive assumption that only pre-defined objects from the source domain exist in the target domain. This closed-set setting prevents the detector from detecting undefined objects. To ease this assumption, we propose Source-Free Unknown Object Detection (SFUOD), a novel scenario which enables the detector to not only recognize known objects but also detect undefined objects as unknown objects. To this end, we propose CollaPAUL (Collaborative tuning and Principal Axis-based Unknown Labeling), a novel framework for SFUOD. Collaborative tuning enhances knowledge adaptation by integrating target-dependent knowledge from the auxiliary encoder with source-dependent knowledge from the pre-trained detector through a cross-domain attention mechanism. Additionally, principal axes-based unknown labeling assigns pseudo-labels to unknown objects by estimating objectness via principal axes projection and confidence scores from model predictions. The proposed CollaPAUL achieves state-of-the-art performances on SFUOD benchmarks, and extensive experiments validate its effectiveness.

View Paper

Investigating Training Data Detection in AI Coders

Authors: Tianlin Li, Yunxiang Wei, Zhiming Li, Aishan Liu, Qing Guo, Xianglong Liu, Dongning Sun, Yang Liu

Abstract: arXiv:2507.17389v1 Announce Type: cross Abstract: Recent advances in code large language models (CodeLLMs) have made them indispensable tools in modern software engineering. However, these models occasionally produce outputs that contain proprietary or sensitive code snippets, raising concerns about potential non-compliant use of training data, and posing risks to privacy and intellectual property. To ensure responsible and compliant deployment of CodeLLMs, training data detection (TDD) has become a critical task. While recent TDD methods have shown promise in natural language settings, their effectiveness on code data remains largely underexplored. This gap is particularly important given code's structured syntax and distinct similarity criteria compared to natural language. To address this, we conduct a comprehensive empirical study of seven state-of-the-art TDD methods on source code data, evaluating their performance across eight CodeLLMs. To support this evaluation, we introduce CodeSnitch, a function-level benchmark dataset comprising 9,000 code samples in three programming languages, each explicitly labeled as either included or excluded from CodeLLM training. Beyond evaluation on the original CodeSnitch, we design targeted mutation strategies to test the robustness of TDD methods under three distinct settings. These mutation strategies are grounded in the well-established Type-1 to Type-4 code clone detection taxonomy. Our study provides a systematic assessment of current TDD techniques for code and offers insights to guide the development of more effective and robust detection methods in the future.

View Paper

HiProbe-VAD: Video Anomaly Detection via Hidden States Probing in Tuning-Free Multimodal LLMs

Authors: Zhaolin Cai, Fan Li, Ziwei Zheng, Yanjun Qin

Abstract: arXiv:2507.17394v1 Announce Type: cross Abstract: Video Anomaly Detection (VAD) aims to identify and locate deviations from normal patterns in video sequences. Traditional methods often struggle with substantial computational demands and a reliance on extensive labeled datasets, thereby restricting their practical applicability. To address these constraints, we propose HiProbe-VAD, a novel framework that leverages pre-trained Multimodal Large Language Models (MLLMs) for VAD without requiring fine-tuning. In this paper, we discover that the intermediate hidden states of MLLMs contain information-rich representations, exhibiting higher sensitivity and linear separability for anomalies compared to the output layer. To capitalize on this, we propose a Dynamic Layer Saliency Probing (DLSP) mechanism that intelligently identifies and extracts the most informative hidden states from the optimal intermediate layer during the MLLMs reasoning. Then a lightweight anomaly scorer and temporal localization module efficiently detects anomalies using these extracted hidden states and finally generate explanations. Experiments on the UCF-Crime and XD-Violence datasets demonstrate that HiProbe-VAD outperforms existing training-free and most traditional approaches. Furthermore, our framework exhibits remarkable cross-model generalization capabilities in different MLLMs without any tuning, unlocking the potential of pre-trained MLLMs for video anomaly detection and paving the way for more practical and scalable solutions.

View Paper

Millions of GeAR-s: Extending GraphRAG to Millions of Documents

Authors: Zhili Shen, Chenxin Diao, Pascual Merita, Pavlos Vougiouklis, Jeff Z. Pan

Abstract: arXiv:2507.17399v1 Announce Type: cross Abstract: Recent studies have explored graph-based approaches to retrieval-augmented generation, leveraging structured or semi-structured information -- such as entities and their relations extracted from documents -- to enhance retrieval. However, these methods are typically designed to address specific tasks, such as multi-hop question answering and query-focused summarisation, and therefore, there is limited evidence of their general applicability across broader datasets. In this paper, we aim to adapt a state-of-the-art graph-based RAG solution: GeAR and explore its performance and limitations on the SIGIR 2025 LiveRAG Challenge.

View Paper

Content-based 3D Image Retrieval and a ColBERT-inspired Re-ranking for Tumor Flagging and Staging

Authors: Farnaz Khun Jush, Steffen Vogler, Matthias Lenga

Abstract: arXiv:2507.17412v1 Announce Type: cross Abstract: The increasing volume of medical images poses challenges for radiologists in retrieving relevant cases. Content-based image retrieval (CBIR) systems offer potential for efficient access to similar cases, yet lack standardized evaluation and comprehensive studies. Building on prior studies for tumor characterization via CBIR, this study advances CBIR research for volumetric medical images through three key contributions: (1) a framework eliminating reliance on pre-segmented data and organ-specific datasets, aligning with large and unstructured image archiving systems, i.e. PACS in clinical practice; (2) introduction of C-MIR, a novel volumetric re-ranking method adapting ColBERT's contextualized late interaction mechanism for 3D medical imaging; (3) comprehensive evaluation across four tumor sites using three feature extractors and three database configurations. Our evaluations highlight the significant advantages of C-MIR. We demonstrate the successful adaptation of the late interaction principle to volumetric medical images, enabling effective context-aware re-ranking. A key finding is C-MIR's ability to effectively localize the region of interest, eliminating the need for pre-segmentation of datasets and offering a computationally efficient alternative to systems relying on expensive data enrichment steps. C-MIR demonstrates promising improvements in tumor flagging, achieving improved performance, particularly for colon and lung tumors ($p<0.05$). C-MIR also shows potential for improving tumor staging, warranting further exploration of its capabilities. Ultimately, our work seeks to bridge the gap between advanced retrieval techniques and their practical applications in healthcare, paving the way for improved diagnostic processes.

View Paper

Fair Compromises in Participatory Budgeting: a Multi-Agent Deep Reinforcement Learning Approach

Authors: Hugh Adams, Srijoni Majumdar, Evangelos Pournaras

Abstract: arXiv:2507.17433v1 Announce Type: cross Abstract: Participatory budgeting is a method of collectively understanding and addressing spending priorities where citizens vote on how a budget is spent, it is regularly run to improve the fairness of the distribution of public funds. Participatory budgeting requires voters to make decisions on projects which can lead to ``choice overload". A multi-agent reinforcement learning approach to decision support can make decision making easier for voters by identifying voting strategies that increase the winning proportion of their vote. This novel approach can also support policymakers by highlighting aspects of election design that enable fair compromise on projects. This paper presents a novel, ethically aligned approach to decision support using multi-agent deep reinforcement learning modelling. This paper introduces a novel use of a branching neural network architecture to overcome scalability challenges of multi-agent reinforcement learning in a decentralized way. Fair compromises are found through optimising voter actions towards greater representation of voter preferences in the winning set. Experimental evaluation with real-world participatory budgeting data reveals a pattern in fair compromise: that it is achievable through projects with smaller cost.

View Paper

Each to Their Own: Exploring the Optimal Embedding in RAG

Authors: Shiting Chen, Zijian Zhao, Jinsong Chen

Abstract: arXiv:2507.17442v1 Announce Type: cross Abstract: Recently, as Large Language Models (LLMs) have fundamentally impacted various fields, the methods for incorporating up-to-date information into LLMs or adding external knowledge to construct domain-specific models have garnered wide attention. Retrieval-Augmented Generation (RAG), serving as an inference-time scaling method, is notable for its low cost and minimal effort for parameter tuning. However, due to heterogeneous training data and model architecture, the variant embedding models used in RAG exhibit different benefits across various areas, often leading to different similarity calculation results and, consequently, varying response quality from LLMs. To address this problem, we propose and examine two approaches to enhance RAG by combining the benefits of multiple embedding models, named Mixture-Embedding RAG and Confident RAG. Mixture-Embedding RAG simply sorts and selects retrievals from multiple embedding models based on standardized similarity; however, it does not outperform vanilla RAG. In contrast, Confident RAG generates responses multiple times using different embedding models and then selects the responses with the highest confidence level, demonstrating average improvements of approximately 10% and 5% over vanilla LLMs and RAG, respectively. The consistent results across different LLMs and embedding models indicate that Confident RAG is an efficient plug-and-play approach for various domains. We will release our code upon publication.

View Paper

IndoorBEV: Joint Detection and Footprint Completion of Objects via Mask-based Prediction in Indoor Scenarios for Bird's-Eye View Perception

Authors: Haichuan Li, Changda Tian, Panos Trahanias, Tomi Westerlund

Abstract: arXiv:2507.17445v1 Announce Type: cross Abstract: Detecting diverse objects within complex indoor 3D point clouds presents significant challenges for robotic perception, particularly with varied object shapes, clutter, and the co-existence of static and dynamic elements where traditional bounding box methods falter. To address these limitations, we propose IndoorBEV, a novel mask-based Bird's-Eye View (BEV) method for indoor mobile robots. In a BEV method, a 3D scene is projected into a 2D BEV grid which handles naturally occlusions and provides a consistent top-down view aiding to distinguish static obstacles from dynamic agents. The obtained 2D BEV results is directly usable to downstream robotic tasks like navigation, motion prediction, and planning. Our architecture utilizes an axis compact encoder and a window-based backbone to extract rich spatial features from this BEV map. A query-based decoder head then employs learned object queries to concurrently predict object classes and instance masks in the BEV space. This mask-centric formulation effectively captures the footprint of both static and dynamic objects regardless of their shape, offering a robust alternative to bounding box regression. We demonstrate the effectiveness of IndoorBEV on a custom indoor dataset featuring diverse object classes including static objects and dynamic elements like robots and miscellaneous items, showcasing its potential for robust indoor scene understanding.

View Paper

Reasoning-Driven Retrosynthesis Prediction with Large Language Models via Reinforcement Learning

Authors: Situo Zhang, Hanqi Li, Lu Chen, Zihan Zhao, Xuanze Lin, Zichen Zhu, Bo Chen, Xin Chen, Kai Yu

Abstract: arXiv:2507.17448v1 Announce Type: cross Abstract: Retrosynthesis planning, essential in organic synthesis and drug discovery, has greatly benefited from recent AI-driven advancements. Nevertheless, existing methods frequently face limitations in both applicability and explainability. Traditional graph-based and sequence-to-sequence models often lack generalized chemical knowledge, leading to predictions that are neither consistently accurate nor easily explainable. To address these challenges, we introduce RetroDFM-R, a reasoning-based large language model (LLM) designed specifically for chemical retrosynthesis. Leveraging large-scale reinforcement learning guided by chemically verifiable rewards, RetroDFM-R significantly enhances prediction accuracy and explainability. Comprehensive evaluations demonstrate that RetroDFM-R significantly outperforms state-of-the-art methods, achieving a top-1 accuracy of 65.0% on the USPTO-50K benchmark. Double-blind human assessments further validate the chemical plausibility and practical utility of RetroDFM-R's predictions. RetroDFM-R also accurately predicts multistep retrosynthetic routes reported in the literature for both real-world drug molecules and perovskite materials. Crucially, the model's explicit reasoning process provides human-interpretable insights, thereby enhancing trust and practical value in real-world retrosynthesis applications.

View Paper

Probing Vision-Language Understanding through the Visual Entailment Task: promises and pitfalls

Authors: Elena Pitta, Tom Kouwenhoven, Tessa Verhoef

Abstract: arXiv:2507.17467v1 Announce Type: cross Abstract: This study investigates the extent to which the Visual Entailment (VE) task serves as a reliable probe of vision-language understanding in multimodal language models, using the LLaMA 3.2 11B Vision model as a test case. Beyond reporting performance metrics, we aim to interpret what these results reveal about the underlying possibilities and limitations of the VE task. We conduct a series of experiments across zero-shot, few-shot, and fine-tuning settings, exploring how factors such as prompt design, the number and order of in-context examples and access to visual information might affect VE performance. To further probe the reasoning processes of the model, we used explanation-based evaluations. Results indicate that three-shot inference outperforms the zero-shot baselines. However, additional examples introduce more noise than they provide benefits. Additionally, the order of the labels in the prompt is a critical factor that influences the predictions. In the absence of visual information, the model has a strong tendency to hallucinate and imagine content, raising questions about the model's over-reliance on linguistic priors. Fine-tuning yields strong results, achieving an accuracy of 83.3% on the e-SNLI-VE dataset and outperforming the state-of-the-art OFA-X model. Additionally, the explanation evaluation demonstrates that the fine-tuned model provides semantically meaningful explanations similar to those of humans, with a BERTScore F1-score of 89.2%. We do, however, find comparable BERTScore results in experiments with limited vision, questioning the visual grounding of this task. Overall, our results highlight both the utility and limitations of VE as a diagnostic task for vision-language understanding and point to directions for refining multimodal evaluation methods.

View Paper

Demonstration of Efficient Predictive Surrogates for Large-scale Quantum Processors

Authors: Wei-You Liao, Yuxuan Du, Xinbiao Wang, Tian-Ci Tian, Yong Luo, Bo Du, Dacheng Tao, He-Liang Huang

Abstract: arXiv:2507.17470v1 Announce Type: cross Abstract: The ongoing development of quantum processors is driving breakthroughs in scientific discovery. Despite this progress, the formidable cost of fabricating large-scale quantum processors means they will remain rare for the foreseeable future, limiting their widespread application. To address this bottleneck, we introduce the concept of predictive surrogates, which are classical learning models designed to emulate the mean-value behavior of a given quantum processor with provably computational efficiency. In particular, we propose two predictive surrogates that can substantially reduce the need for quantum processor access in diverse practical scenarios. To demonstrate their potential in advancing digital quantum simulation, we use these surrogates to emulate a quantum processor with up to 20 programmable superconducting qubits, enabling efficient pre-training of variational quantum eigensolvers for families of transverse-field Ising models and identification of non-equilibrium Floquet symmetry-protected topological phases. Experimental results reveal that the predictive surrogates not only reduce measurement overhead by orders of magnitude, but can also surpass the performance of conventional, quantum-resource-intensive approaches. Collectively, these findings establish predictive surrogates as a practical pathway to broadening the impact of advanced quantum processors.

View Paper

BGM-HAN: A Hierarchical Attention Network for Accurate and Fair Decision Assessment on Semi-Structured Profiles

Authors: Junhua Liu, Roy Ka-Wei Lee, Kwan Hui Lim

Abstract: arXiv:2507.17472v1 Announce Type: cross Abstract: Human decision-making in high-stakes domains often relies on expertise and heuristics, but is vulnerable to hard-to-detect cognitive biases that threaten fairness and long-term outcomes. This work presents a novel approach to enhancing complex decision-making workflows through the integration of hierarchical learning alongside various enhancements. Focusing on university admissions as a representative high-stakes domain, we propose BGM-HAN, an enhanced Byte-Pair Encoded, Gated Multi-head Hierarchical Attention Network, designed to effectively model semi-structured applicant data. BGM-HAN captures multi-level representations that are crucial for nuanced assessment, improving both interpretability and predictive performance. Experimental results on real admissions data demonstrate that our proposed model significantly outperforms both state-of-the-art baselines from traditional machine learning to large language models, offering a promising framework for augmenting decision-making in domains where structure, context, and fairness matter. Source code is available at: https://github.com/junhua/bgm-han.

View Paper

MultiNRC: A Challenging and Native Multilingual Reasoning Evaluation Benchmark for LLMs

Authors: Alexander R. Fabbri, Diego Mares, Jorge Flores, Meher Mankikar, Ernesto Hernandez, Dean Lee, Bing Liu, Chen Xing

Abstract: arXiv:2507.17476v1 Announce Type: cross Abstract: Although recent Large Language Models (LLMs) have shown rapid improvement on reasoning benchmarks in English, the evaluation of such LLMs' multilingual reasoning capability across diverse languages and cultural contexts remains limited. Existing multilingual reasoning benchmarks are typically constructed by translating existing English reasoning benchmarks, biasing these benchmarks towards reasoning problems with context in English language/cultures. In this work, we introduce the Multilingual Native Reasoning Challenge (MultiNRC), a benchmark designed to assess LLMs on more than 1,000 native, linguistic and culturally grounded reasoning questions written by native speakers in French, Spanish, and Chinese. MultiNRC covers four core reasoning categories: language-specific linguistic reasoning, wordplay & riddles, cultural/tradition reasoning, and math reasoning with cultural relevance. For cultural/tradition reasoning and math reasoning with cultural relevance, we also provide English equivalent translations of the multilingual questions by manual translation from native speakers fluent in English. This set of English equivalents can provide a direct comparison of LLM reasoning capacity in other languages vs. English on the same reasoning questions. We systematically evaluate current 14 leading LLMs covering most LLM families on MultiNRC and its English equivalent set. The results show that (1) current LLMs are still not good at native multilingual reasoning, with none scoring above 50% on MultiNRC; (2) LLMs exhibit distinct strengths and weaknesses in handling linguistic, cultural, and logical reasoning tasks; (3) Most models perform substantially better in math reasoning in English compared to in original languages (+10%), indicating persistent challenges with culturally grounded knowledge.

View Paper

Unsupervised anomaly detection using Bayesian flow networks: application to brain FDG PET in the context of Alzheimer's disease

Authors: Hugues Roy, Reuben Dorent, Ninon Burgos

Abstract: arXiv:2507.17486v1 Announce Type: cross Abstract: Unsupervised anomaly detection (UAD) plays a crucial role in neuroimaging for identifying deviations from healthy subject data and thus facilitating the diagnosis of neurological disorders. In this work, we focus on Bayesian flow networks (BFNs), a novel class of generative models, which have not yet been applied to medical imaging or anomaly detection. BFNs combine the strength of diffusion frameworks and Bayesian inference. We introduce AnoBFN, an extension of BFNs for UAD, designed to: i) perform conditional image generation under high levels of spatially correlated noise, and ii) preserve subject specificity by incorporating a recursive feedback from the input image throughout the generative process. We evaluate AnoBFN on the challenging task of Alzheimer's disease-related anomaly detection in FDG PET images. Our approach outperforms other state-of-the-art methods based on VAEs (beta-VAE), GANs (f-AnoGAN), and diffusion models (AnoDDPM), demonstrating its effectiveness at detecting anomalies while reducing false positive rates.

View Paper

To Trust or Not to Trust: On Calibration in ML-based Resource Allocation for Wireless Networks

Authors: Rashika Raina, Nidhi Simmons, David E. Simmons, Michel Daoud Yacoub, Trung Q. Duong

Abstract: arXiv:2507.17494v1 Announce Type: cross Abstract: In next-generation communications and networks, machine learning (ML) models are expected to deliver not only accurate predictions but also well-calibrated confidence scores that reflect the true likelihood of correct decisions. This paper studies the calibration performance of an ML-based outage predictor within a single-user, multi-resource allocation framework. We first establish key theoretical properties of this system's outage probability (OP) under perfect calibration. Importantly, we show that as the number of resources grows, the OP of a perfectly calibrated predictor approaches the expected output conditioned on it being below the classification threshold. In contrast, when only one resource is available, the system's OP equals the model's overall expected output. We then derive the OP conditions for a perfectly calibrated predictor. These findings guide the choice of the classification threshold to achieve a desired OP, helping system designers meet specific reliability requirements. We also demonstrate that post-processing calibration cannot improve the system's minimum achievable OP, as it does not introduce new information about future channel states. Additionally, we show that well-calibrated models are part of a broader class of predictors that necessarily improve OP. In particular, we establish a monotonicity condition that the accuracy-confidence function must satisfy for such improvement to occur. To demonstrate these theoretical properties, we conduct a rigorous simulation-based analysis using post-processing calibration techniques: Platt scaling and isotonic regression. As part of this framework, the predictor is trained using an outage loss function specifically designed for this system. Furthermore, this analysis is performed on Rayleigh fading channels with temporal correlation captured by Clarke's 2D model, which accounts for receiver mobility.

View Paper

HOTA: Hamiltonian framework for Optimal Transport Advection

Authors: Nazar Buzun, Daniil Shlenskii, Maxim Bobrin, Dmitry V. Dylov

Abstract: arXiv:2507.17513v1 Announce Type: cross Abstract: Optimal transport (OT) has become a natural framework for guiding the probability flows. Yet, the majority of recent generative models assume trivial geometry (e.g., Euclidean) and rely on strong density-estimation assumptions, yielding trajectories that do not respect the true principles of optimality in the underlying manifold. We present Hamiltonian Optimal Transport Advection (HOTA), a Hamilton-Jacobi-Bellman based method that tackles the dual dynamical OT problem explicitly through Kantorovich potentials, enabling efficient and scalable trajectory optimization. Our approach effectively evades the need for explicit density modeling, performing even when the cost functionals are non-smooth. Empirically, HOTA outperforms all baselines in standard benchmarks, as well as in custom datasets with non-differentiable costs, both in terms of feasibility and optimality.

View Paper

Enabling Cyber Security Education through Digital Twins and Generative AI

Authors: Vita Santa Barletta, Vito Bavaro, Miriana Calvano, Antonio Curci, Antonio Piccinno, Davide Pio Posa

Abstract: arXiv:2507.17518v1 Announce Type: cross Abstract: Digital Twins (DTs) are gaining prominence in cybersecurity for their ability to replicate complex IT (Information Technology), OT (Operational Technology), and IoT (Internet of Things) infrastructures, allowing for real time monitoring, threat analysis, and system simulation. This study investigates how integrating DTs with penetration testing tools and Large Language Models (LLMs) can enhance cybersecurity education and operational readiness. By simulating realistic cyber environments, this approach offers a practical, interactive framework for exploring vulnerabilities and defensive strategies. At the core of this research is the Red Team Knife (RTK), a custom penetration testing toolkit aligned with the Cyber Kill Chain model. RTK is designed to guide learners through key phases of cyberattacks, including reconnaissance, exploitation, and response within a DT powered ecosystem. The incorporation of Large Language Models (LLMs) further enriches the experience by providing intelligent, real-time feedback, natural language threat explanations, and adaptive learning support during training exercises. This combined DT LLM framework is currently being piloted in academic settings to develop hands on skills in vulnerability assessment, threat detection, and security operations. Initial findings suggest that the integration significantly improves the effectiveness and relevance of cybersecurity training, bridging the gap between theoretical knowledge and real-world application. Ultimately, the research demonstrates how DTs and LLMs together can transform cybersecurity education to meet evolving industry demands.

View Paper

Integrating Physics-Based and Data-Driven Approaches for Probabilistic Building Energy Modeling

Authors: Leandro Von Krannichfeldt, Kristina Orehounig, Olga Fink

Abstract: arXiv:2507.17526v1 Announce Type: cross Abstract: Building energy modeling is a key tool for optimizing the performance of building energy systems. Historically, a wide spectrum of methods has been explored -- ranging from conventional physics-based models to purely data-driven techniques. Recently, hybrid approaches that combine the strengths of both paradigms have gained attention. These include strategies such as learning surrogates for physics-based models, modeling residuals between simulated and observed data, fine-tuning surrogates with real-world measurements, using physics-based outputs as additional inputs for data-driven models, and integrating the physics-based output into the loss function the data-driven model. Despite this progress, two significant research gaps remain. First, most hybrid methods focus on deterministic modeling, often neglecting the inherent uncertainties caused by factors like weather fluctuations and occupant behavior. Second, there has been little systematic comparison within a probabilistic modeling framework. This study addresses these gaps by evaluating five representative hybrid approaches for probabilistic building energy modeling, focusing on quantile predictions of building thermodynamics in a real-world case study. Our results highlight two main findings. First, the performance of hybrid approaches varies across different building room types, but residual learning with a Feedforward Neural Network performs best on average. Notably, the residual approach is the only model that produces physically intuitive predictions when applied to out-of-distribution test data. Second, Quantile Conformal Prediction is an effective procedure for calibrating quantile predictions in case of indoor temperature modeling.

View Paper

Federated Majorize-Minimization: Beyond Parameter Aggregation

Authors: Aymeric Dieuleveut, Gersende Fort, Mahmoud Hegazy, Hoi-To Wai

Abstract: arXiv:2507.17534v1 Announce Type: cross Abstract: This paper proposes a unified approach for designing stochastic optimization algorithms that robustly scale to the federated learning setting. Our work studies a class of Majorize-Minimization (MM) problems, which possesses a linearly parameterized family of majorizing surrogate functions. This framework encompasses (proximal) gradient-based algorithms for (regularized) smooth objectives, the Expectation Maximization algorithm, and many problems seen as variational surrogate MM. We show that our framework motivates a unifying algorithm called Stochastic Approximation Stochastic Surrogate MM (\SSMM), which includes previous stochastic MM procedures as special instances. We then extend \SSMM\ to the federated setting, while taking into consideration common bottlenecks such as data heterogeneity, partial participation, and communication constraints; this yields \QSMM. The originality of \QSMM\ is to learn locally and then aggregate information characterizing the \textit{surrogate majorizing function}, contrary to classical algorithms which learn and aggregate the \textit{original parameter}. Finally, to showcase the flexibility of this methodology beyond our theoretical setting, we use it to design an algorithm for computing optimal transport maps in the federated setting.

View Paper

Enhancing Quantum Federated Learning with Fisher Information-Based Optimization

Authors: Amandeep Singh Bhatia, Sabre Kais

Abstract: arXiv:2507.17580v1 Announce Type: cross Abstract: Federated Learning (FL) has become increasingly popular across different sectors, offering a way for clients to work together to train a global model without sharing sensitive data. It involves multiple rounds of communication between the global model and participating clients, which introduces several challenges like high communication costs, heterogeneous client data, prolonged processing times, and increased vulnerability to privacy threats. In recent years, the convergence of federated learning and parameterized quantum circuits has sparked significant research interest, with promising implications for fields such as healthcare and finance. By enabling decentralized training of quantum models, it allows clients or institutions to collaboratively enhance model performance and outcomes while preserving data privacy. Recognizing that Fisher information can quantify the amount of information that a quantum state carries under parameter changes, thereby providing insight into its geometric and statistical properties. We intend to leverage this property to address the aforementioned challenges. In this work, we propose a Quantum Federated Learning (QFL) algorithm that makes use of the Fisher information computed on local client models, with data distributed across heterogeneous partitions. This approach identifies the critical parameters that significantly influence the quantum model's performance, ensuring they are preserved during the aggregation process. Our research assessed the effectiveness and feasibility of QFL by comparing its performance against other variants, and exploring the benefits of incorporating Fisher information in QFL settings. Experimental results on ADNI and MNIST datasets demonstrate the effectiveness of our approach in achieving better performance and robustness against the quantum federated averaging method.

View Paper

PRIX: Learning to Plan from Raw Pixels for End-to-End Autonomous Driving

Authors: Maciej K. Wozniak, Lianhang Liu, Yixi Cai, Patric Jensfelt

Abstract: arXiv:2507.17596v1 Announce Type: cross Abstract: While end-to-end autonomous driving models show promising results, their practical deployment is often hindered by large model sizes, a reliance on expensive LiDAR sensors and computationally intensive BEV feature representations. This limits their scalability, especially for mass-market vehicles equipped only with cameras. To address these challenges, we propose PRIX (Plan from Raw Pixels). Our novel and efficient end-to-end driving architecture operates using only camera data, without explicit BEV representation and forgoing the need for LiDAR. PRIX leverages a visual feature extractor coupled with a generative planning head to predict safe trajectories from raw pixel inputs directly. A core component of our architecture is the Context-aware Recalibration Transformer (CaRT), a novel module designed to effectively enhance multi-level visual features for more robust planning. We demonstrate through comprehensive experiments that PRIX achieves state-of-the-art performance on the NavSim and nuScenes benchmarks, matching the capabilities of larger, multimodal diffusion planners while being significantly more efficient in terms of inference speed and model size, making it a practical solution for real-world deployment. Our work is open-source and the code will be at https://maxiuw.github.io/prix.

View Paper

Vision Transformer attention alignment with human visual perception in aesthetic object evaluation

Authors: Miguel Carrasco, C\'esar Gonz\'alez-Mart\'in, Jos\'e Aranda, Luis Oliveros

Abstract: arXiv:2507.17616v1 Announce Type: cross Abstract: Visual attention mechanisms play a crucial role in human perception and aesthetic evaluation. Recent advances in Vision Transformers (ViTs) have demonstrated remarkable capabilities in computer vision tasks, yet their alignment with human visual attention patterns remains underexplored, particularly in aesthetic contexts. This study investigates the correlation between human visual attention and ViT attention mechanisms when evaluating handcrafted objects. We conducted an eye-tracking experiment with 30 participants (9 female, 21 male, mean age 24.6 years) who viewed 20 artisanal objects comprising basketry bags and ginger jars. Using a Pupil Labs eye-tracker, we recorded gaze patterns and generated heat maps representing human visual attention. Simultaneously, we analyzed the same objects using a pre-trained ViT model with DINO (Self-DIstillation with NO Labels), extracting attention maps from each of the 12 attention heads. We compared human and ViT attention distributions using Kullback-Leibler divergence across varying Gaussian parameters (sigma=0.1 to 3.0). Statistical analysis revealed optimal correlation at sigma=2.4 +-0.03, with attention head #12 showing the strongest alignment with human visual patterns. Significant differences were found between attention heads, with heads #7 and #9 demonstrating the greatest divergence from human attention ($p< 0.05$, Tukey HSD test). Results indicate that while ViTs exhibit more global attention patterns compared to human focal attention, certain attention heads can approximate human visual behavior, particularly for specific object features like buckles in basketry items. These findings suggest potential applications of ViT attention mechanisms in product design and aesthetic evaluation, while highlighting fundamental differences in attention strategies between human perception and current AI models.

View Paper

How Should We Meta-Learn Reinforcement Learning Algorithms?

Authors: Alexander David Goldie, Zilin Wang, Jakob Nicolaus Foerster, Shimon Whiteson

Abstract: arXiv:2507.17668v1 Announce Type: cross Abstract: The process of meta-learning algorithms from data, instead of relying on manual design, is growing in popularity as a paradigm for improving the performance of machine learning systems. Meta-learning shows particular promise for reinforcement learning (RL), where algorithms are often adapted from supervised or unsupervised learning despite their suboptimality for RL. However, until now there has been a severe lack of comparison between different meta-learning algorithms, such as using evolution to optimise over black-box functions or LLMs to propose code. In this paper, we carry out this empirical comparison of the different approaches when applied to a range of meta-learned algorithms which target different parts of the RL pipeline. In addition to meta-train and meta-test performance, we also investigate factors including the interpretability, sample cost and train time for each meta-learning algorithm. Based on these findings, we propose several guidelines for meta-learning new RL algorithms which will help ensure that future learned algorithms are as performant as possible.

View Paper

CASCADE: LLM-Powered JavaScript Deobfuscator at Google

Authors: Shan Jiang, Pranoy Kovuri, David Tao, Zhixun Tan

Abstract: arXiv:2507.17691v1 Announce Type: cross Abstract: Software obfuscation, particularly prevalent in JavaScript, hinders code comprehension and analysis, posing significant challenges to software testing, static analysis, and malware detection. This paper introduces CASCADE, a novel hybrid approach that integrates the advanced coding capabilities of Gemini with the deterministic transformation capabilities of a compiler Intermediate Representation (IR), specifically JavaScript IR (JSIR). By employing Gemini to identify critical prelude functions, the foundational components underlying the most prevalent obfuscation techniques, and leveraging JSIR for subsequent code transformations, CASCADE effectively recovers semantic elements like original strings and API names, and reveals original program behaviors. This method overcomes limitations of existing static and dynamic deobfuscation techniques, eliminating hundreds to thousands of hardcoded rules while achieving reliability and flexibility. CASCADE is already deployed in Google's production environment, demonstrating substantial improvements in JavaScript deobfuscation efficiency and reducing reverse engineering efforts.

View Paper

From Feedback to Checklists: Grounded Evaluation of AI-Generated Clinical Notes

Authors: Karen Zhou, John Giorgi, Pranav Mani, Peng Xu, Davis Liang, Chenhao Tan

Abstract: arXiv:2507.17717v1 Announce Type: cross Abstract: AI-generated clinical notes are increasingly used in healthcare, but evaluating their quality remains a challenge due to high subjectivity and limited scalability of expert review. Existing automated metrics often fail to align with real-world physician preferences. To address this, we propose a pipeline that systematically distills real user feedback into structured checklists for note evaluation. These checklists are designed to be interpretable, grounded in human feedback, and enforceable by LLM-based evaluators. Using deidentified data from over 21,000 clinical encounters, prepared in accordance with the HIPAA safe harbor standard, from a deployed AI medical scribe system, we show that our feedback-derived checklist outperforms baseline approaches in our offline evaluations in coverage, diversity, and predictive power for human ratings. Extensive experiments confirm the checklist's robustness to quality-degrading perturbations, significant alignment with clinician preferences, and practical value as an evaluation methodology. In offline research settings, the checklist can help identify notes likely to fall below our chosen quality thresholds.

View Paper

AI Telephone Surveying: Automating Quantitative Data Collection with an AI Interviewer

Authors: Danny D. Leybzon, Shreyas Tirumala, Nishant Jain, Summer Gillen, Michael Jackson, Cameron McPhee, Jennifer Schmidt

Abstract: arXiv:2507.17718v1 Announce Type: cross Abstract: With the rise of voice-enabled artificial intelligence (AI) systems, quantitative survey researchers have access to a new data-collection mode: AI telephone surveying. By using AI to conduct phone interviews, researchers can scale quantitative studies while balancing the dual goals of human-like interactivity and methodological rigor. Unlike earlier efforts that used interactive voice response (IVR) technology to automate these surveys, voice AI enables a more natural and adaptive respondent experience as it is more robust to interruptions, corrections, and other idiosyncrasies of human speech. We built and tested an AI system to conduct quantitative surveys based on large language models (LLM), automatic speech recognition (ASR), and speech synthesis technologies. The system was specifically designed for quantitative research, and strictly adhered to research best practices like question order randomization, answer order randomization, and exact wording. To validate the system's effectiveness, we deployed it to conduct two pilot surveys with the SSRS Opinion Panel and followed-up with a separate human-administered survey to assess respondent experiences. We measured three key metrics: the survey completion rates, break-off rates, and respondent satisfaction scores. Our results suggest that shorter instruments and more responsive AI interviewers may contribute to improvements across all three metrics studied.

View Paper

On the Interaction of Compressibility and Adversarial Robustness

Authors: Melih Barsbey, Ant\^onio H. Ribeiro, Umut \c{S}im\c{s}ekli, Tolga Birdal

Abstract: arXiv:2507.17725v1 Announce Type: cross Abstract: Modern neural networks are expected to simultaneously satisfy a host of desirable properties:

accurate fitting to training data, generalization to unseen inputs, parameter and computational efficiency, and robustness to adversarial perturbations. While compressibility and robustness have each been studied extensively, a unified understanding of their interaction still remains elusive. In this work, we develop a principled framework to analyze how different forms of compressibility - such as neuron-level sparsity and spectral compressibility - affect adversarial robustness. We show that these forms of compression can induce a small number of highly sensitive directions in the representation space, which adversaries can exploit to construct effective perturbations. Our analysis yields a simple yet instructive robustness bound, revealing how neuron and spectral compressibility impact $L|*infty$ and $L*2$ robustness via their effects on the learned representations. Crucially, the vulnerabilities we identify arise irrespective of how compression is achieved - whether via regularization, architectural bias, or implicit learning dynamics. Through empirical evaluations across synthetic and realistic tasks, we confirm our theoretical predictions, and further demonstrate that these vulnerabilities persist under adversarial training and transfer learning, and contribute to the emergence of universal adversarial perturbations. Our findings show a fundamental tension between structured compressibility and robustness, and suggest new pathways for designing models that are both efficient and secure.

View Paper

Flow Matching Meets Biology and Life Science: A Survey

Authors: Zihao Li, Zhichen Zeng, Xiao Lin, Feihao Fang, Yanru Qu, Zhe Xu, Zhining Liu, Xuying Ning, Tianxin Wei, Ge Liu, Hanghang Tong, Jingrui He

Abstract: arXiv:2507.17731v1 Announce Type: cross Abstract: Over the past decade, advances in generative modeling, such as generative adversarial networks, masked autoencoders, and diffusion models, have significantly transformed biological research and discovery, enabling breakthroughs in molecule design, protein generation, drug discovery, and beyond. At the same time, biological applications have served as valuable testbeds for evaluating the capabilities of generative models. Recently, flow matching has emerged as a powerful and efficient alternative to diffusion-based generative modeling, with growing interest in its application to problems in biology and life sciences. This paper presents the first comprehensive survey of recent developments in flow matching and its applications in biological domains. We begin by systematically reviewing the foundations and variants of flow matching, and then categorize its applications into three major areas: biological sequence modeling, molecule generation and design, and peptide and protein generation. For each, we provide an in-depth review of recent progress. We also summarize commonly used datasets and software tools, and conclude with a discussion of potential future directions. The corresponding curated resources are available at https://github.com/Violet24K/Awesome-Flow-Matching-Meets-Biology.

View Paper

Yume: An Interactive World Generation Model

Authors: Xiaofeng Mao, Shaoheng Lin, Zhen Li, Chuanhao Li, Wenshuo Peng, Tong He, Jiangmiao Pang, Mingmin Chi, Yu Qiao, Kaipeng Zhang

Abstract: arXiv:2507.17744v1 Announce Type: cross Abstract: Yume aims to use images, text, or videos to create an interactive, realistic, and dynamic world, which allows exploration and control using peripheral devices or neural signals. In this report, we present a preview version of \method, which creates a dynamic world from an input image and allows exploration of the world using keyboard actions. To achieve this high-fidelity and interactive video world generation, we introduce a well-designed framework, which consists of four main components, including camera motion quantization, video generation architecture, advanced sampler, and model acceleration. First, we quantize camera motions for stable training and user-friendly interaction using keyboard inputs. Then, we introduce the Masked Video Diffusion Transformer~(MVDT) with a memory module for infinite video generation in an autoregressive manner. After that, training-free Anti-Artifact Mechanism (AAM) and Time Travel Sampling based on Stochastic Differential Equations (TTS-SDE) are introduced to the sampler for better visual quality and more precise control. Moreover, we investigate model acceleration by synergistic optimization of adversarial distillation and caching mechanisms. We use the high-quality world exploration dataset \sekai to train \method, and it achieves remarkable results in diverse scenes and applications. All data, codebase, and model weights are available on https://github.com/stdstu12/YUME. Yume will update monthly to achieve its original goal. Project page: https://stdstu12.github.io/YUME-Project/.

View Paper

Ultra3D: Efficient and High-Fidelity 3D Generation with Part Attention

Authors: Yiwen Chen, Zhihao Li, Yikai Wang, Hu Zhang, Qin Li, Chi Zhang, Guosheng Lin

Abstract: arXiv:2507.17745v1 Announce Type: cross Abstract: Recent advances in sparse voxel representations have significantly improved the quality of 3D content generation, enabling high-resolution modeling with fine-grained geometry. However, existing frameworks suffer from severe computational inefficiencies due to the quadratic complexity of attention mechanisms in their two-stage diffusion pipelines. In this work, we propose Ultra3D, an efficient 3D generation framework that significantly accelerates sparse voxel modeling without compromising quality. Our method leverages the compact VecSet representation to efficiently generate a coarse object layout in the first stage, reducing token count and accelerating voxel coordinate prediction. To refine per-voxel latent features in the second stage, we introduce Part Attention, a geometry-aware localized attention mechanism that restricts attention computation within semantically consistent part regions. This design preserves structural continuity while avoiding unnecessary global attention, achieving up to 6.7x speed-up in latent generation. To support this mechanism, we construct a scalable part annotation pipeline that converts raw meshes into part-labeled sparse voxels. Extensive experiments demonstrate that Ultra3D supports high-resolution 3D generation at 1024 resolution and achieves state-of-the-art performance in both visual fidelity and user preference.

View Paper

Rubrics as Rewards: Reinforcement Learning Beyond Verifiable Domains

Authors: Anisha Gunjal, Anthony Wang, Elaine Lau, Vaskar Nath, Bing Liu, Sean Hendryx

Abstract: arXiv:2507.17746v1 Announce Type: cross Abstract: Extending Reinforcement Learning with Verifiable Rewards (RLVR) to real-world tasks often requires balancing objective and subjective evaluation criteria. However, many such tasks lack a single, unambiguous ground truth-making it difficult to define reliable reward signals for post-training language models. While traditional preference-based methods offer a workaround, they rely on opaque reward functions that are difficult to interpret and prone to spurious correlations. We introduce $\textbf{Rubrics as Rewards}$ (RaR), a framework that uses structured, checklist-style rubrics as interpretable reward signals for on-policy training with GRPO. Our best RaR method yields up to a 28% relative improvement on HealthBench-1k compared to simple Likert-based approaches, while matching or surpassing the performance of reward signals derived from expert-written references. By treating rubrics as structured reward signals, we show that RaR enables smaller-scale judge models to better align with human preferences and sustain robust performance across model scales.

View Paper

Pretraining on the Test Set Is No Longer All You Need: A Debate-Driven Approach to QA Benchmarks

Authors: Linbo Cao, Jinman Zhao

Abstract: arXiv:2507.17747v1 Announce Type: cross Abstract: As frontier language models increasingly saturate standard QA benchmarks, concerns about data contamination, memorization, and escalating dataset creation costs persist. We propose a debate-driven evaluation paradigm that transforms any existing QA dataset into structured adversarial debates--where one model is given the official answer to defend, and another constructs and defends an alternative answer--adjudicated by a judge model blind to the correct solution. By forcing multi-round argumentation, this approach substantially increases difficulty while penalizing shallow memorization, yet reuses QA items to reduce curation overhead. We make two main contributions: (1) an evaluation pipeline to systematically convert QA tasks into debate-based assessments, and (2) a public benchmark that demonstrates our paradigm's effectiveness on a subset of MMLU-Pro questions, complete with standardized protocols and reference models. Empirical results validate the robustness of the method and its effectiveness against data contamination--a Llama 3.1 model fine-tuned on test questions showed dramatic accuracy improvements (50% -> 82%) but performed worse in debates. Results also show that even weaker judges can reliably differentiate stronger debaters, highlighting how debate-based evaluation can scale to future, more capable systems while maintaining a fraction of the cost of creating new benchmarks. Overall, our framework underscores that "pretraining on the test set is no longer all you need," offering a sustainable path for measuring the genuine reasoning ability of advanced language models.

View Paper

Large Learning Rates Simultaneously Achieve Robustness to Spurious Correlations and Compressibility

Authors: Melih Barsbey, Lucas Prieto, Stefanos Zafeiriou, Tolga Birdal

Abstract: arXiv:2507.17748v1 Announce Type: cross Abstract: Robustness and resource-efficiency are two highly desirable properties for modern machine learning models. However, achieving them jointly remains a challenge. In this paper, we position high learning rates as a facilitator for simultaneously achieving robustness to spurious correlations and network compressibility. We demonstrate that large learning rates also produce desirable representation properties such as invariant feature utilization, class separation, and activation sparsity. Importantly, our findings indicate that large learning rates compare favorably to other hyperparameters and regularization methods, in consistently satisfying these properties in tandem. In addition to demonstrating the positive effect of large learning rates across diverse spurious correlation datasets, models, and optimizers, we also present strong evidence that the previously documented success of large learning rates in standard classification tasks is likely due to its effect on addressing hidden/rare spurious correlations in the training dataset.

View Paper

Conflict Detection for Temporal Knowledge Graphs:A Fast Constraint Mining Algorithm and New Benchmarks

Authors: Jianhao Chen, Junyang Ren, Wentao Ding, Haoyuan Ouyang, Wei Hu, Yuzhong Qu

Abstract: arXiv:2312.11053v2 Announce Type: replace Abstract: Temporal facts, which are used to describe events that occur during specific time periods, have become a topic of increased interest in the field of knowledge graph (KG) research. In terms of quality management, the introduction of time restrictions brings new challenges to maintaining the temporal consistency of KGs. Previous studies rely on manually enumerated temporal constraints to detect conflicts, which are labor-intensive and may have granularity issues. To address this problem, we start from the common pattern of temporal facts and propose a pattern-based temporal constraint mining method, PaTeCon. Unlike previous studies, PaTeCon uses graph patterns and statistical information relevant to the given KG to automatically generate temporal constraints, without the need for human experts. In this paper, we illustrate how this method can be optimized to achieve significant speed improvement. We also annotate Wikidata and Freebase to build two new benchmarks for conflict detection. Extensive experiments demonstrate that our pattern-based automatic constraint mining approach is highly effective in generating valuable temporal constraints.

View Paper

LLM as a code generator in Agile Model Driven Development

Authors: Ahmed R. Sadik, Sebastian Brulin, Markus Olhofer

Abstract: arXiv:2410.18489v2 Announce Type: replace Abstract: Leveraging Large Language Models (LLM) like GPT4 in the auto generation of code represents a significant advancement, yet it is not without its challenges. The ambiguity inherent in natural language descriptions of software poses substantial obstacles to generating deployable, structured artifacts. This research champions Model Driven Development (MDD) as a viable strategy to overcome these challenges, proposing an Agile Model Driven Development (AMDD) approach that employs GPT4 as a code generator. This approach enhances the flexibility and scalability of the code auto generation process and offers agility that allows seamless adaptation to changes in models or deployment environments. We illustrate this by modeling a multi agent Unmanned Vehicle Fleet (UVF) system using the Unified Modeling Language (UML), significantly reducing model ambiguity by integrating the Object Constraint Language (OCL) for code structure meta modeling, and the FIPA ontology language for communication semantics meta modeling. Applying GPT4 auto generation capabilities yields Java and Python code that is compatible with the JADE and PADE frameworks, respectively. Our thorough evaluation of the auto generated code verifies its alignment with expected behaviors and identifies enhancements in agent interactions. Structurally, we assessed the complexity of code derived from a model constrained solely by OCL meta models, against that influenced by both OCL and FIPA ontology meta models. The results indicate that the ontology constrained meta model produces inherently more complex code, yet its cyclomatic complexity remains within manageable levels, suggesting that additional meta model constraints can be incorporated without exceeding the high risk threshold for complexity.

View Paper

Learning Neural Strategy-Proof Matching Mechanism from Examples

Authors: Ryota Maruo, Koh Takeuchi, Hisashi Kashima

Abstract: arXiv:2410.19384v2 Announce Type: replace Abstract: Designing two-sided matching mechanisms is challenging when practical demands for matching outcomes are difficult to formalize and the designed mechanism must satisfy theoretical conditions. To address this, prior work has proposed a framework that learns a matching mechanism from examples, using a parameterized family that satisfies properties such as stability. However, despite its usefulness, this framework does not guarantee strategy-proofness (SP), and cannot handle varying numbers of agents or incorporate publicly available contextual information about agents, both of which are crucial in real-world applications. In this paper, we propose a new parametrized family of matching mechanisms that always satisfy strategy-proofness, are applicable for an arbitrary number of

agents, and deal with public contextual information of agents, based on the serial dictatorship (SD). This family is represented by NeuralSD, a novel neural network architecture based on SD, where agent rankings in SD are treated as learnable parameters computed from agents' contexts using an attention-based sub-network. To enable learning, we introduce tensor serial dictatorship (TSD), a differentiable relaxation of SD using tensor operations. This allows NeuralSD to be trained end-to-end from example matchings while satisfying SP. We conducted experiments to learn a matching mechanism from matching examples while satisfying SP. We demonstrated that our method outperformed baselines in predicting matchings and on several metrics for goodness of matching outcomes.

View Paper

Balans: Multi-Armed Bandits-based Adaptive Large Neighborhood Search for Mixed-Integer Programming Problem

Authors: Junyang Cai, Serdar Kadioglu, Bistra Dilkina

Abstract: arXiv:2412.14382v3 Announce Type: replace Abstract: Mixed-integer programming (MIP) is a powerful paradigm for modeling and solving various important combinatorial optimization problems. Recently, learning-based approaches have shown a potential to speed up MIP solving via offline training that then guides important design decisions during the search. However, a significant drawback of these methods is their heavy reliance on offline training, which requires collecting training datasets and computationally costly training epochs yet offering only limited generalization to unseen (larger) instances. In this paper, we propose Balans, an adaptive meta-solver for MIPs with online learning capability that does not require any supervision or apriori training. At its core, Balans is based on adaptive large-neighborhood search, operating on top of an MIP solver by successive applications of destroy and repair neighborhood operators. During the search, the selection among different neighborhood definitions is guided on the fly for the instance at hand via multi-armed bandit algorithms. Our extensive experiments on hard optimization instances show that Balans offers significant performance gains over the default MIP solver, is better than committing to any single best neighborhood, and improves over the state-of-the-art large-neighborhood search for MIPs. Finally, we release Balans as a highly configurable, MIP solver agnostic, open-source software.

View Paper

A novel approach to navigate the taxonomic hierarchy to address the Open-World Scenarios in Medicinal Plant Classification

Authors: Soumen Sinha, Tanisha Rana, Rahul Roy

Abstract: arXiv:2502.17289v3 Announce Type: replace Abstract: In this article, we propose a novel approach for plant hierarchical taxonomy classification by posing the problem as an open class problem. It is observed that existing methods for medicinal plant classification often fail to perform hierarchical classification and accurately identifying unknown species, limiting their effectiveness in comprehensive plant taxonomy classification. Thus we address the problem of unknown species classification by assigning it best hierarchical labels. We propose a novel method, which integrates DenseNet121, Multi-Scale Self-Attention (MSSA) and cascaded classifiers for hierarchical classification. The approach systematically categorizes medicinal plants at multiple taxonomic levels, from phylum to species, ensuring detailed and precise classification. Using multi scale space attention, the model captures both local and global contextual information from the images, improving the distinction between similar species and the identification of new ones. It uses attention scores to focus on important features across multiple scales. The proposed method provides a solution for hierarchical classification, showcasing superior performance in identifying both known and unknown species. The model was tested on two state-of-art datasets with and without background artifacts and so that it can be deployed to tackle real word application. We used unknown species for testing our model. For unknown species the model achieved an average accuracy of 83.36%, 78.30%, 60.34% and 43.32% for predicting correct phylum, class, order and family respectively. Our proposed model size is almost four times less than the existing state of the art methods making it easily deploy able in real world application.

View Paper

From Hypothesis to Publication: A Comprehensive Survey of AI-Driven Research Support Systems

Authors: Zekun Zhou, Xiaocheng Feng, Lei Huang, Xiachong Feng, Ziyun Song, Ruihan Chen, Liang Zhao, Weitao Ma, Yuxuan Gu, Baoxin Wang, Dayong Wu, Guoping Hu, Ting Liu, Bing Qin

Abstract: arXiv:2503.01424v2 Announce Type: replace Abstract: Research is a fundamental process driving the advancement of human civilization, yet it demands substantial time and effort from researchers. In recent years, the rapid development of artificial intelligence (AI) technologies has inspired researchers to explore how AI can accelerate and enhance research. To monitor relevant advancements, this paper presents a systematic review of the progress in this domain. Specifically, we organize the relevant studies into three main categories: hypothesis formulation, hypothesis validation, and manuscript publication. Hypothesis formulation involves knowledge synthesis and hypothesis generation. Hypothesis validation includes the verification of scientific claims, theorem proving, and experiment validation. Manuscript publication encompasses manuscript writing and the peer review process. Furthermore, we identify and discuss the current challenges faced in these areas, as well as potential future directions for research. Finally, we also offer a comprehensive overview of existing benchmarks and tools across various domains that support the integration of AI into the research process. We hope this paper serves as an introduction for beginners and fosters future research. Resources have been made publicly available at https://github.com/zkzhou126/AI-for-Research.

View Paper

More is Less: The Pitfalls of Multi-Model Synthetic Preference Data in DPO Safety Alignment

Authors: Yifan Wang, Runjin Chen, Bolian Li, David Cho, Yihe Deng, Ruqi Zhang, Tianlong Chen, Zhangyang Wang, Ananth Grama, Junyuan Hong

Abstract: arXiv:2504.02193v2 Announce Type: replace Abstract: Aligning large language models (LLMs) with human values is an increasingly critical step in post-training. Direct Preference Optimization (DPO) has emerged as a simple, yet effective alternative to reinforcement learning from human feedback (RLHF). Synthetic preference data with its low cost and high quality enable effective alignment through single- or multi-model generated preference data. Our study reveals a striking, safety-specific phenomenon associated with DPO alignment: Although multi-model generated data enhances performance on general tasks (ARC, Hellaswag, MMLU, TruthfulQA, Winogrande) by providing diverse responses, it also tends to facilitate reward hacking during training. This can lead to a high attack success rate (ASR) when models encounter jailbreaking prompts. The issue is particularly pronounced when employing stronger models like GPT-4o or larger models in the same family to generate chosen responses paired with target model self-generated rejected responses, resulting in dramatically poorer safety outcomes. Furthermore, with respect to safety, using solely self-generated responses (single-model generation) for both chosen and rejected pairs significantly outperforms configurations that incorporate responses from stronger models, whether used directly as chosen data or as part of a multi-model response pool. We demonstrate that multi-model preference data exhibits high linear separability between chosen and rejected responses, which allows models to exploit superficial cues rather than internalizing robust safety constraints. Our experiments, conducted on models from the Llama, Mistral, and Qwen families, consistently validate these findings.

View Paper

Agent RL Scaling Law: Agent RL with Spontaneous Code Execution for Mathematical Problem Solving

Authors: Xinji Mai, Haotian Xu, Xing W, Weinong Wang, Jian Hu, Yingying Zhang, Wenqiang Zhang

Abstract: arXiv:2505.07773v3 Announce Type: replace Abstract: Large Language Models (LLMs) often struggle with mathematical reasoning tasks requiring precise, verifiable computation. While Reinforcement Learning (RL) from outcome-based rewards enhances text-based reasoning, understanding how agents autonomously learn to leverage external tools like code execution remains crucial. We investigate RL from outcome-based rewards for Tool-Integrated Reasoning, ZeroTIR, training base LLMs to spontaneously generate and execute Python code for mathematical problems without supervised tool-use examples. Our central contribution is we demonstrate that as RL training progresses, key metrics scale predictably. Specifically, we observe strong positive correlations where increased training steps lead to increases in the spontaneous code execution frequency, the average response length, and, critically, the final task accuracy. This suggests a quantifiable relationship between computational effort invested in training and the emergence of effective, tool-augmented reasoning strategies. We implement a robust framework featuring a decoupled code execution environment and validate our findings across standard RL algorithms and frameworks. Experiments show ZeroTIR significantly surpasses non-tool ZeroRL baselines on challenging math benchmarks. Our findings provide a foundational understanding of how autonomous tool use is acquired and scales within Agent RL, offering a reproducible benchmark for future studies. Code is released at \href{https://github.com/yyht/openrlhf*async*pipline}{https://github.com/yyht/openrlhf_async_pipline}.

View Paper

Turing Test 2.0: The General Intelligence Threshold

Authors: Georgios Mappouras

Abstract: arXiv:2505.19550v4 Announce Type: replace Abstract: With the rise of artificial intelligence (A.I.) and large language models like ChatGPT, a new race for achieving artificial general intelligence (A.G.I) has started. While many speculate how and when A.I. will achieve A.G.I., there is no clear agreement on how A.G.I. can be detected in A.I. models, even when popular tools like the Turing test (and its modern variations) are used to measure their intelligence. In this work, we discuss why traditional methods like the Turing test do not suffice for measuring or detecting A.G.I. and provide a new, practical method that can be used to decide if a system (computer or any other) has reached or surpassed A.G.I. To achieve this, we make two new contributions. First, we present a clear definition for general intelligence (G.I.) and set a G.I. Threshold (G.I.T.) that can be used to distinguish between systems that achieve A.G.I. and systems that do not. Second, we present a new framework on how to construct tests that can detect if a system has achieved G.I. in a simple, comprehensive, and clear-cut fail/pass way. We call this novel framework the Turing test 2.0. We then demonstrate real-life examples of applying tests that follow our Turing test 2.0 framework on modern A.I. models.

View Paper

Tournament of Prompts: Evolving LLM Instructions Through Structured Debates and Elo Ratings

Authors: Anirudh Nair, Adi Banerjee, Laurent Mombaerts, Matthew Hagen, Tarik Borogovac

Abstract: arXiv:2506.00178v2 Announce Type: replace Abstract: Prompt engineering represents a critical bottleneck to harness the full potential of Large Language Models (LLMs) for solving complex tasks, as it requires specialized expertise, significant trial-and-error, and manual intervention. This challenge is particularly pronounced for tasks involving subjective quality assessment, where defining explicit optimization objectives becomes fundamentally problematic. Existing automated prompt optimization methods falter in these scenarios, as they typically require well-defined task-specific numerical fitness functions or rely on generic templates that cannot capture the nuanced requirements of complex use cases. We introduce DEEVO (DEbate-driven EVOlutionary prompt optimization), a novel framework that guides prompt evolution through a debate-driven evaluation with an Elo-based selection. Contrary to prior work, DEEVOs approach enables exploration of the discrete prompt space while preserving semantic coherence through intelligent crossover and strategic mutation operations that incorporate debate-based feedback, combining elements from both successful and unsuccessful prompts based on identified strengths rather than arbitrary splicing. Using Elo ratings as a fitness proxy, DEEVO simultaneously drives improvement and preserves valuable diversity in the prompt population. Experimental results demonstrate that DEEVO significantly outperforms both manual prompt engineering and alternative state-of-the-art optimization approaches on open-ended tasks and close-ended tasks despite using no ground truth feedback. By connecting LLMs reasoning capabilities with adaptive optimization, DEEVO represents a significant advancement in prompt optimization research by eliminating the need of predetermined metrics to continuously improve AI systems.

View Paper

A Community-driven vision for a new Knowledge Resource for AI

Authors: Vinay K Chaudhri, Chaitan Baru, Brandon Bennett, Mehul Bhatt, Darion Cassel, Anthony G Cohn, Rina Dechter, Esra Erdem, Dave Ferrucci, Ken Forbus, Gregory Gelfond, Michael Genesereth, Andrew S. Gordon, Benjamin Grosof, Gopal Gupta, Jim Hendler, Sharat Israni, Tyler R. Josephson, Patrick Kyllonen, Yuliya Lierler, Vladimir Lifschitz, Clifton McFate, Hande K. McGinty, Leora Morgenstern, Alessandro Oltramari, Praveen Paritosh, Dan Roth, Blake Shepard, Cogan Shimzu, Denny Vrande\v{c}i\'c, Mark Whiting, Michael Witbrock

Abstract: arXiv:2506.16596v2 Announce Type: replace Abstract: The long-standing goal of creating a comprehensive, multi-purpose knowledge resource, reminiscent of the 1984 Cyc project, still persists in AI. Despite the success of knowledge resources like WordNet, ConceptNet, Wolfram|Alpha and other commercial knowledge graphs, verifiable, general-purpose widely available sources of knowledge remain a critical deficiency in AI infrastructure. Large language models struggle due to knowledge gaps; robotic planning lacks necessary world knowledge; and the detection of factually false information relies heavily on human expertise. What kind of knowledge resource is most needed in AI today? How can modern technology shape its development and evaluation? A recent AAAI workshop gathered over 50 researchers to explore these questions. This paper synthesizes our findings and outlines a community-driven vision for a new knowledge infrastructure. In addition to leveraging contemporary advances in knowledge representation and reasoning, one promising idea is to build an open engineering framework to exploit knowledge modules effectively within the context of practical applications. Such a framework should include sets of conventions and social structures that are adopted by contributors.

View Paper

LEGO Co-builder: Exploring Fine-Grained Vision-Language Modeling for Multimodal LEGO Assembly Assistants

Authors: Haochen Huang, Jiahuan Pei, Mohammad Aliannejadi, Xin Sun, Moonisa Ahsan, Chuang Yu, Zhaochun Ren, Pablo Cesar, Junxiao Wang

Abstract: arXiv:2507.05515v2 Announce Type: replace Abstract: Vision-language models (VLMs) are facing the challenges of understanding and following multimodal assembly instructions, particularly when fine-grained spatial reasoning and precise object state detection are required. In this work, we explore LEGO Co-builder, a hybrid benchmark combining real-world LEGO assembly logic with programmatically generated multimodal scenes. The dataset captures stepwise visual states and procedural instructions, allowing controlled evaluation of instruction-following, object detection, and state detection. We introduce a unified framework and assess leading VLMs such as GPT-4o, Gemini, and Qwen-VL, under zero-shot and fine-tuned settings. Our results reveal that even advanced models like GPT-4o struggle with fine-grained assembly tasks, with a maximum F1 score of just 40.54\% on state detection, highlighting gaps in fine-grained visual understanding. We release the benchmark, codebase, and generation pipeline to support future research on multimodal assembly assistants grounded in real-world workflows.

View Paper

Working with AI: Measuring the Occupational Implications of Generative AI

Authors: Kiran Tomlinson, Sonia Jaffe, Will Wang, Scott Counts, Siddharth Suri

Abstract: arXiv:2507.07935v3 Announce Type: replace Abstract: Given the rapid adoption of generative AI and its potential to impact a wide range of tasks, understanding the effects of AI on the economy is one of society's most important questions. In this work, we take a step toward that goal by analyzing the work activities people do with AI, how successfully and broadly those activities are done, and combine that with data on what occupations do those activities. We analyze a dataset of 200k anonymized and privacy-scrubbed conversations between users and Microsoft Bing Copilot, a publicly available generative AI system. We find the most common work activities people seek AI assistance for involve gathering information and writing, while the most common activities that AI itself is performing are providing information and assistance, writing, teaching, and advising. Combining these activity classifications with measurements of task success and scope of impact, we compute an AI applicability score for each occupation. We find the highest AI applicability scores for knowledge work occupation groups such as computer and mathematical, and office and administrative support, as well as occupations such as sales whose work activities involve providing and communicating information. Additionally, we characterize the types of work activities performed most successfully, how wage and education correlate with AI applicability, and how real-world usage compares to predictions of occupational AI impact.

View Paper

Automated planning with ontologies under coherence update semantics (Extended Version)

Authors: Stefan Borgwardt, Duy Nhu, Gabriele R\"oger

Abstract: arXiv:2507.15120v2 Announce Type: replace Abstract: Standard automated planning employs first-order formulas under closed-world semantics to achieve a goal with a given set of actions from an initial state. We follow a line of research that aims to incorporate background knowledge into automated planning problems, for example, by means of ontologies, which are usually interpreted under open-world semantics. We present a new approach for planning with DL-Lite ontologies that combines the advantages of ontology-based action conditions provided by explicit-input knowledge and action bases (eKABs) and ontology-aware action effects under the coherence update semantics. We show that the complexity of the resulting formalism is not higher than that of previous approaches and provide an implementation via a polynomial compilation into classical planning. An evaluation of existing and new benchmarks examines the performance of a planning system on different variants of our compilation.

View Paper

Deliberative Searcher: Improving LLM Reliability via Reinforcement Learning with constraints

Authors: Zhenyun Yin, Shujie Wang, Xuhong Wang, Xingjun Ma, Yinchun Wang

Abstract: arXiv:2507.16727v2 Announce Type: replace Abstract: Improving the reliability of large language models (LLMs) is critical for deploying them in real-world scenarios. In this paper, we propose \textbf{Deliberative Searcher}, the first framework to integrate certainty calibration with retrieval-based search for open-domain question answering. The agent performs multi-step reflection and verification over Wikipedia data and is trained with a reinforcement learning algorithm that optimizes for accuracy under a soft reliability constraint.

Empirical results show that proposed method improves alignment between model confidence and correctness, leading to more trustworthy outputs. This paper will be continuously updated.

View Paper

ACMP: Allen-Cahn Message Passing with Attractive and Repulsive Forces for Graph Neural Networks

Authors: Yuelin Wang, Kai Yi, Xinliang Liu, Yu Guang Wang, Shi Jin

Abstract: arXiv:2206.05437v4 Announce Type: replace-cross Abstract: Neural message passing is a basic feature extraction unit for graph-structured data considering neighboring node features in network propagation from one layer to the next. We model such process by an interacting particle system with attractive and repulsive forces and the Allen-Cahn force arising in the modeling of phase transition. The dynamics of the system is a reaction-diffusion process which can separate particles without blowing up. This induces an Allen-Cahn message passing (ACMP) for graph neural networks where the numerical iteration for the particle system solution constitutes the message passing propagation. ACMP which has a simple implementation with a neural ODE solver can propel the network depth up to one hundred of layers with theoretically proven strictly positive lower bound of the Dirichlet energy. It thus provides a deep model of GNNs circumventing the common GNN problem of oversmoothing. GNNs with ACMP achieve state of the art performance for real-world node classification tasks on both homophilic and heterophilic datasets. Codes are available at https://github.com/ykiiiiii/ACMP.

View Paper

From DDMs to DNNs: Using process data and models of decision-making to improve human-AI interactions

Authors: Mrugsen Nagsen Gopnarayan, Jaan Aru, Sebastian Gluth

Abstract: arXiv:2308.15225v3 Announce Type: replace-cross Abstract: Over the past decades, cognitive neuroscientists and behavioral economists have recognized the value of describing the process of decision making in detail and modeling the emergence of decisions over time. For example, the time it takes to decide can reveal more about an agent's true hidden preferences than only the decision itself. Similarly, data that track the ongoing decision process such as eye movements or neural recordings contain critical information that can be exploited, even if no decision is made. Here, we argue that artificial intelligence (AI) research would benefit from a stronger focus on insights about how decisions emerge over time and incorporate related process data to improve AI predictions in general and human-AI interactions in particular. First, we introduce a highly established computational framework that assumes decisions to emerge from the noisy accumulation of evidence, and we present related empirical work in psychology, neuroscience, and economics. Next, we discuss to what extent current approaches in multi-agent AI do or do not incorporate process data and models of decision making. Finally, we outline how a more principled inclusion of the evidence-accumulation framework into the training and use of AI can help to improve human-AI interactions in the future.

View Paper

Towards Efficient Generative Large Language Model Serving: A Survey from Algorithms to Systems

Authors: Xupeng Miao, Gabriele Oliaro, Zhihao Zhang, Xinhao Cheng, Hongyi Jin, Tianqi Chen, Zhihao Jia

Abstract: arXiv:2312.15234v2 Announce Type: replace-cross Abstract: In the rapidly evolving landscape of artificial intelligence (AI), generative large language models (LLMs) stand at the forefront, revolutionizing how we interact with our data. However, the computational intensity and memory consumption of deploying these models present substantial challenges in terms of serving efficiency, particularly in scenarios demanding low latency and high throughput. This survey addresses the imperative need for efficient LLM serving methodologies from a machine learning system (MLSys) research perspective, standing at the crux of advanced AI innovations and practical system optimizations. We provide in-depth analysis, covering a spectrum of solutions, ranging from cutting-edge algorithmic modifications to groundbreaking changes in system designs. The survey aims to provide a comprehensive understanding of the current state and future directions in efficient LLM serving, offering valuable insights for researchers and practitioners in overcoming the barriers of effective LLM deployment, thereby reshaping the future of AI.

View Paper

Enhancing Sequential Recommender with Large Language Models for Joint Video and Comment Recommendation

Authors: Bowen Zheng, Zihan Lin, Enze Liu, Chen Yang, Enyang Bai, Cheng Ling, Wayne Xin Zhao, Ji-Rong Wen

Abstract: arXiv:2403.13574v2 Announce Type: replace-cross Abstract: Nowadays, reading or writing comments on captivating videos has emerged as a critical part of the viewing experience on online video platforms. However, existing recommender systems primarily focus on users' interaction behaviors with videos, neglecting comment content and interaction in user preference modeling. In this paper, we propose a novel recommendation approach called LSVCR that utilizes user interaction histories with both videos and comments to jointly perform personalized video and comment recommendation. Specifically, our approach comprises two key components: sequential recommendation (SR) model and supplemental large language model (LLM) recommender. The SR model functions as the primary recommendation backbone (retained in deployment) of our method for efficient user preference modeling. Concurrently, we employ a LLM as the supplemental recommender (discarded in deployment) to better capture underlying user preferences derived from heterogeneous interaction behaviors. In order to integrate the strengths of the SR model and the supplemental LLM recommender, we introduce a two-stage training paradigm. The first stage, personalized preference alignment, aims to align the preference representations from both components, thereby enhancing the semantics of the SR model. The second stage, recommendation-oriented fine-tuning, involves fine-tuning the alignment-enhanced SR model according to specific objectives. Extensive experiments in both video and comment recommendation tasks demonstrate the effectiveness of LSVCR. Moreover, online A/B testing on KuaiShou platform verifies the practical benefits of our approach. In particular, we attain a cumulative gain of 4.13% in comment watch time.

View Paper

Multi-Level Explanations for Generative Language Models

Authors: Lucas Monteiro Paes, Dennis Wei, Hyo Jin Do, Hendrik Strobelt, Ronny Luss, Amit Dhurandhar, Manish Nagireddy, Karthikeyan Natesan Ramamurthy, Prasanna Sattigeri, Werner Geyer, Soumya Ghosh

Abstract: arXiv:2403.14459v2 Announce Type: replace-cross Abstract: Despite the increasing use of large language models (LLMs) for context-grounded tasks like

summarization and question-answering, understanding what makes an LLM produce a certain response is challenging. We propose Multi-Level Explanations for Generative Language Models (MExGen), a technique to provide explanations for context-grounded text generation. MExGen assigns scores to parts of the context to quantify their influence on the model's output. It extends attribution methods like LIME and SHAP to LLMs used in context-grounded tasks where (1) inference cost is high, (2) input text is long, and (3) the output is text. We conduct a systematic evaluation, both automated and human, of perturbation-based attribution methods for summarization and question answering. The results show that our framework can provide more faithful explanations of generated output than available alternatives, including LLM self-explanations. We open-source code for MExGen as part of the ICX360 toolkit: https://github$.$com/IBM/ICX360.

View Paper

Constructing Optimal Noise Channels for Enhanced Robustness in Quantum Machine Learning

Authors: David Winderl, Nicola Franco, Jeanette Miriam Lorenz

Abstract: arXiv:2404.16417v2 Announce Type: replace-cross Abstract: With the rapid advancement of Quantum Machine Learning (QML), the critical need to enhance security measures against adversarial attacks and protect QML models becomes increasingly evident. In this work, we outline the connection between quantum noise channels and differential privacy (DP), by constructing a family of noise channels which are inherently ϵ-DP: (α, γ)-channels. Through this approach, we successfully replicate the ϵ-DP bounds observed for depolarizing and random rotation channels, thereby affirming the broad generality of our framework. Additionally, we use a semi-definite program to construct an optimally robust channel. In a small-scale experimental evaluation, we demonstrate the benefits of using our optimal noise channel over depolarizing noise, particularly in enhancing adversarial accuracy. Moreover, we assess how the variables α and γ affect the certifiable robustness and investigate how different encoding methods impact the classifier's robustness.

View Paper

Impact of Stickers on Multimodal Sentiment and Intent in Social Media: A New Task, Dataset and Baseline

Authors: Yuanchen Shi, Biao Ma, Longyin Zhang, Fang Kong

Abstract: arXiv:2405.08427v2 Announce Type: replace-cross Abstract: Stickers are increasingly used in social media to express sentiment and intent. Despite their significant impact on sentiment analysis and intent recognition, little research has been conducted in this area. To address this gap, we propose a new task: \textbf{M}ultimodal chat \textbf{S}entiment \textbf{A}nalysis and \textbf{I}ntent \textbf{R}ecognition involving \textbf{S}tickers (MSAIRS). Additionally, we introduce a novel multimodal dataset containing Chinese chat records and stickers excerpted from several mainstream social media platforms. Our dataset includes paired data with the same text but different stickers, the same sticker but different contexts, and various stickers consisting of the same images with different texts, allowing us to better understand the impact of stickers on chat sentiment and intent. We also propose an effective multimodal joint model, MMSAIR, featuring differential vector construction and cascaded attention mechanisms for enhanced multimodal fusion. Our experiments demonstrate the necessity and effectiveness of jointly modeling sentiment and intent, as they mutually reinforce each other's recognition accuracy. MMSAIR significantly outperforms traditional models and advanced MLLMs, demonstrating the challenge and uniqueness of sticker interpretation in social media. Our dataset and code are available on https://github.com/FakerBoom/MSAIRS-Dataset.

View Paper

Cross-domain Multi-step Thinking: Zero-shot Fine-grained Traffic Sign Recognition in the Wild

Authors: Yaozong Gan, Guang Li, Ren Togo, Keisuke Maeda, Takahiro Ogawa, Miki Haseyama

Abstract: arXiv:2409.01534v2 Announce Type: replace-cross Abstract: In this study, we propose Cross-domain Multi-step Thinking (CdMT) to improve zero-shot fine-grained traffic sign recognition (TSR) performance in the wild. Zero-shot fine-grained TSR in the wild is challenging due to the cross-domain problem between clean template traffic signs and real-world counterparts, and existing approaches particularly struggle with cross-country TSR scenarios, where traffic signs typically differ between countries. The proposed CdMT framework tackles these challenges by leveraging the multi-step reasoning capabilities of large multimodal models (LMMs). We introduce context, characteristic, and differential descriptions to design multiple thinking processes for LMMs. Context descriptions, which are enhanced by center coordinate prompt optimization, enable the precise localization of target traffic signs in complex road images and filter irrelevant responses via novel prior traffic sign hypotheses. Characteristic descriptions, which are derived from in-context learning with template traffic signs, bridge cross-domain gaps and enhance fine-grained TSR. Differential descriptions refine the multimodal reasoning ability of LMMs by distinguishing subtle differences among similar signs. CdMT is independent of training data and requires only simple and uniform instructions, enabling it to achieve cross-country TSR. We conducted extensive experiments on three benchmark datasets and two real-world datasets from different countries. The proposed CdMT framework achieved superior performance compared with other state-of-the-art methods on all five datasets, with recognition accuracies of 0.93, 0.89, 0.97, 0.89, and 0.85 on the GTSRB, BTSD, TT-100K, Sapporo, and Yokohama datasets, respectively.

View Paper

The FIX Benchmark: Extracting Features Interpretable to eXperts

Authors: Helen Jin, Shreya Havaldar, Chaehyeon Kim, Anton Xue, Weiqiu You, Helen Qu, Marco Gatti, Daniel A Hashimoto, Bhuvnesh Jain, Amin Madani, Masao Sako, Lyle Ungar, Eric Wong

Abstract: arXiv:2409.13684v4 Announce Type: replace-cross Abstract: Feature-based methods are commonly used to explain model predictions, but these methods often implicitly assume that interpretable features are readily available. However, this is often not the case for high-dimensional data, and it can be hard even for domain experts to mathematically specify which features are important. Can we instead automatically extract collections or groups of features that are aligned with expert knowledge? To address this gap, we present FIX (Features Interpretable to eXperts), a benchmark for measuring how well a collection of features aligns with expert knowledge. In collaboration with domain experts, we propose FIXScore, a unified expert alignment measure applicable to diverse real-world settings across cosmology, psychology, and medicine domains in vision, language, and time series data modalities. With FIXScore, we find that popular feature-based explanation methods have poor alignment with expert-specified knowledge, highlighting the need for new methods that can better identify features interpretable to experts.

View Paper

BadHMP: Backdoor Attack against Human Motion Prediction

Authors: Chaohui Xu, Si Wang, Chip-Hong Chang

Abstract: arXiv:2409.19638v2 Announce Type: replace-cross Abstract: Precise future human motion prediction over sub-second horizons from past observations is crucial for various safety-critical applications. To date, only a few studies have examined the vulnerability of skeleton-based neural networks to evasion and backdoor attacks. In this paper, we propose BadHMP, a novel backdoor attack that targets specifically human motion prediction tasks. Our approach involves generating poisoned training samples by embedding a localized backdoor trigger in one limb of the skeleton, causing selected joints to follow predefined motion in historical time steps. Subsequently, the future sequences are globally modified that all the joints move following the target trajectories. Our carefully designed backdoor triggers and targets guarantee the smoothness and naturalness of the poisoned samples, making them stealthy enough to evade detection by the model trainer while keeping the poisoned model unobtrusive in terms of prediction fidelity to untainted sequences. The target sequences can be successfully activated by the designed input sequences even with a low poisoned sample injection ratio. Experimental results on two datasets (Human3.6M and CMU-Mocap) and two network architectures (LTD and HRI) demonstrate the high-fidelity, effectiveness, and stealthiness of BadHMP. Robustness of our attack against fine-tuning defense is also verified.

View Paper

Language model developers should report train-test overlap

Authors: Andy K Zhang, Kevin Klyman, Yifan Mai, Yoav Levine, Yian Zhang, Rishi Bommasani, Percy Liang

Abstract: arXiv:2410.08385v2 Announce Type: replace-cross Abstract: Language models are extensively evaluated, but correctly interpreting evaluation results requires knowledge of train-test overlap which refers to the extent to which the language model is trained on the very data it is being tested on. The public currently lacks adequate information about train-test overlap: most models have no public train-test overlap statistics, and third parties cannot directly measure train-test overlap since they do not have access to the training data. To make this clear, we document the practices of 30 model developers, finding that just 9 developers report train-test overlap: 4 developers release training data under open-source licenses, enabling the community to directly measure train-test overlap, and 5 developers publish their train-test overlap methodology and statistics. By engaging with language model developers, we provide novel information about train-test overlap for three additional developers. Overall, we take the position that language model developers should publish train-test overlap statistics and/or training data whenever they report evaluation results on public test sets. We hope our work increases transparency into train-test overlap to increase the community-wide trust in model evaluations.

View Paper

Vascular Segmentation of Functional Ultrasound Images using Deep Learning

Authors: Hana Sebia (AISTROSIGHT), Thomas Guyet (AISTROSIGHT), Micka\"el Pereira (CERMEP - imagerie du vivant), Marco Valdebenito (CERMEP - imagerie du vivant), Hugues Berry (AISTROSIGHT), Benjamin Vidal (CERMEP - imagerie du vivant, CRNL, UCBL)

Abstract: arXiv:2410.22365v2 Announce Type: replace-cross Abstract: Segmentation of medical images is a fundamental task with numerous applications. While MRI, CT, and PET modalities have significantly benefited from deep learning segmentation techniques, more recent modalities, like functional ultrasound (fUS), have seen limited progress. fUS is a non invasive imaging method that measures changes in cerebral blood volume (CBV) with high spatio-temporal resolution. However, distinguishing arterioles from venules in fUS is challenging due to opposing blood flow directions within the same pixel. Ultrasound localization microscopy (ULM) can enhance resolution by tracking microbubble contrast agents but is invasive, and lacks dynamic CBV quantification. In this paper, we introduce the first deep learning-based segmentation tool for fUS images, capable of differentiating signals from different vascular compartments, based on ULM automatic annotation and enabling dynamic CBV quantification. We evaluate various UNet architectures on fUS images of rat brains, achieving competitive segmentation performance, with 90% accuracy, a 71% F1 score, and an IoU of 0.59, using only 100 temporal frames from a fUS stack. These results are comparable to those from tubular structure segmentation in other imaging modalities. Additionally, models trained on resting-state data generalize well to images captured during visual stimulation, highlighting robustness. This work offers a non-invasive, cost-effective alternative to ULM, enhancing fUS data interpretation and improving understanding of vessel function. Our pipeline shows high linear correlation coefficients between signals from predicted and actual compartments in both cortical and deeper regions, showcasing its ability to accurately capture blood flow dynamics.

View Paper

Flexible Coded Distributed Convolution Computing for Enhanced Straggler Resilience and Numerical Stability in Distributed CNNs

Authors: Shuo Tan, Rui Liu, Xuesong Han, XianLei Long, Kai Wan, Linqi Song, Yong Li

Abstract: arXiv:2411.01579v2 Announce Type: replace-cross Abstract: Deploying Convolutional Neural Networks (CNNs) on resource-constrained devices necessitates efficient management of computational resources, often via distributed environments susceptible to latency from straggler nodes. This paper introduces the Flexible Coded Distributed Convolution Computing (FCDCC) framework to enhance straggler resilience and numerical stability in distributed CNNs. We extend Coded Distributed Computing (CDC) with Circulant and Rotation Matrix Embedding (CRME) which was originally proposed for matrix multiplication to high-dimensional tensor convolution. For the proposed scheme, referred to as the Numerically Stable Coded Tensor Convolution (NSCTC) scheme, we also propose two new coded partitioning schemes: Adaptive-Padding Coded Partitioning (APCP) for the input tensor and Kernel-Channel Coded Partitioning (KCCP) for the filter tensor. These strategies enable linear decomposition of tensor convolutions and encoding them into CDC subtasks, combining model parallelism with coded redundancy for robust and efficient execution. Theoretical analysis identifies an optimal trade-off between communication and storage costs. Empirical results validate the framework's effectiveness in computational efficiency, straggler resilience, and scalability across various CNN architectures.

View Paper

A Survey of Event Causality Identification: Taxonomy, Challenges, Assessment, and Prospects

Authors: Qing Cheng, Zefan Zeng, Xingchen Hu, Yuehang Si, Zhong Liu

Abstract: arXiv:2411.10371v4 Announce Type: replace-cross Abstract: Event Causality Identification (ECI) has emerged as a pivotal task in natural language processing (NLP), aimed at automatically detecting causal relationships between events in text. In this comprehensive survey, we systematically elucidate the foundational principles and technical frameworks of ECI, proposing a novel classification framework to categorize and clarify existing methods. {We discuss associated challenges, provide quantitative evaluations, and outline future directions for this dynamic and rapidly evolving field. We first delineate key definitions, problem formalization, and evaluation protocols of ECI. Our classification framework organizes ECI methods based on two primary tasks: Sentence-level Event Causality Identification (SECI) and Document-level Event Causality Identification (DECI). For SECI, we review methods including feature pattern-based matching, machine learning-based classification, deep semantic encoding, prompt-based fine-tuning, and causal knowledge pre-training, alongside common data augmentation strategies. For DECI, we focus on techniques such as deep semantic encoding, event graph reasoning, and prompt-based fine-tuning. We dedicate specific discussions to advancements in multi-lingual and cross-lingual ECI as well as zero-shot ECI leveraging Large Language Models (LLMs). Furthermore, we analyze the strengths, limitations, and unresolved challenges of each method. Extensive quantitative evaluations are conducted on four benchmark datasets to assess various ECI methods. Finally, we explore future research directions.

View Paper

Onto-LLM-TAMP: Knowledge-oriented Task and Motion Planning using Large Language Models

Authors: Muhayy Ud Din, Jan Rosell, Waseem Akram, Isiah Zaplana, Maximo A Roa, Irfan Hussain

Abstract: arXiv:2412.07493v2 Announce Type: replace-cross Abstract: Performing complex manipulation tasks in dynamic environments requires efficient Task and Motion Planning (TAMP) approaches that combine high-level symbolic plans with low-level motion control. Advances in Large Language Models (LLMs), such as GPT-4, are transforming task planning by offering natural language as an intuitive and flexible way to describe tasks, generate symbolic plans, and reason. However, the effectiveness of LLM-based TAMP approaches is limited due to static and template-based prompting, which limits adaptability to dynamic environments and complex task contexts. To address these limitations, this work proposes a novel Onto-LLM-TAMP framework that employs knowledge-based reasoning to refine and expand user prompts with task-contextual reasoning and knowledge-based environment state descriptions. Integrating domain-specific knowledge into the prompt ensures semantically accurate and context-aware task plans. The proposed framework demonstrates its effectiveness by resolving semantic errors in symbolic plan generation, such as maintaining logical temporal goal ordering in scenarios involving hierarchical object placement. The proposed framework is validated through both simulation and real-world scenarios, demonstrating significant improvements over the baseline approach in terms of adaptability to dynamic environments and the generation of semantically correct task plans.

View Paper

NVS-SQA: Exploring Self-Supervised Quality Representation Learning for Neurally Synthesized Scenes without References

Authors: Qiang Qu, Yiran Shen, Xiaoming Chen, Yuk Ying Chung, Weidong Cai, Tongliang Liu

Abstract: arXiv:2501.06488v2 Announce Type: replace-cross Abstract: Neural View Synthesis (NVS), such as NeRF and 3D Gaussian Splatting, effectively creates photorealistic scenes from sparse viewpoints, typically evaluated by quality assessment methods like PSNR, SSIM, and LPIPS. However, these full-reference methods, which compare synthesized views to reference views, may not fully capture the perceptual quality of neurally synthesized scenes (NSS), particularly due to the limited availability of dense reference views. Furthermore, the challenges in acquiring human perceptual labels hinder the creation of extensive labeled datasets, risking model overfitting and reduced generalizability. To address these issues, we propose NVS-SQA, a NSS quality assessment method to learn no-reference quality representations through self-supervision without reliance on human labels. Traditional self-supervised learning predominantly relies on the "same instance, similar representation" assumption and extensive datasets. However, given that these conditions do not apply in NSS quality assessment, we employ heuristic cues and quality scores as learning objectives, along with a specialized contrastive pair preparation process to improve the effectiveness and efficiency of learning. The results show that NVS-SQA outperforms 17 no-reference methods by a large margin (i.e., on average 109.5% in SRCC, 98.6% in PLCC, and 91.5% in KRCC over the second best) and even exceeds 16 full-reference methods across all evaluation metrics (i.e., 22.9% in SRCC, 19.1% in PLCC, and 18.6% in KRCC over the second best).

View Paper

Alleviating Seasickness through Brain-Computer Interface-based Attention Shift

Authors: Xiaoyu Bao, Kailin Xu, Jiawei Zhu, Haiyun Huang, Kangning Li, Qiyun Huang, Yuanqing Li

Abstract: arXiv:2501.08518v2 Announce Type: replace-cross Abstract: Seasickness poses a widespread problem that adversely impacts both passenger comfort and the operational efficiency of maritime crews. Although attention shift has been proposed as a potential method to alleviate symptoms of motion sickness, its efficacy remains to be rigorously validated, especially in maritime environments. In this study, we develop an AI-driven brain-computer interface (BCI) to realize sustained and practical attention shift by incorporating tasks such as breath counting. Forty-three participants completed a real-world nautical experiment consisting of a real-feedback session, a resting session, and a pseudo-feedback session. Notably, 81.39\% of the participants reported that the BCI intervention was effective. EEG analysis revealed that the proposed system can effectively regulate motion sickness EEG signatures, such as an decrease in total band power, along with an increase in theta relative power and a decrease in beta relative power. Furthermore, an indicator of attentional focus, the theta/beta ratio, exhibited a significant reduction during the real-feedback session, providing further evidence to support the effectiveness of the BCI in shifting attention. Collectively, this study presents a novel nonpharmacological, portable, and effective approach for seasickness intervention, which has the potential to open up a brand-new application domain for BCIs.

View Paper

RALAD: Bridging the Real-to-Sim Domain Gap in Autonomous Driving with Retrieval-Augmented Learning

Authors: Jiacheng Zuo, Haibo Hu, Zikang Zhou, Yufei Cui, Ziquan Liu, Jianping Wang, Nan Guan, Jin Wang, Chun Jason Xue

Abstract: arXiv:2501.12296v3 Announce Type: replace-cross Abstract: In the pursuit of robust autonomous driving systems, models trained on real-world datasets often struggle to adapt to new environments, particularly when confronted with corner cases such as extreme weather conditions. Collecting these corner cases in the real world is non-trivial, which necessitates the use of simulators for validation. However,the high computational cost and the domain gap in data distribution have hindered the seamless transition between real and simulated driving scenarios. To tackle this challenge, we propose Retrieval-Augmented Learning for Autonomous Driving (RALAD), a novel framework designed to bridge the real-to-sim gap at a low cost. RALAD features three primary designs, including (1) domain adaptation via an enhanced Optimal Transport (OT) method that accounts for both individual and grouped image distances, (2) a simple and unified framework that can be applied to various models, and (3) efficient fine-tuning techniques that freeze the computationally expensive layers while maintaining robustness. Experimental results demonstrate that RALAD compensates for the performance degradation in simulated environments while maintaining accuracy in real-world scenarios across three different models. Taking Cross View as an example, the mIOU and mAP metrics in real-world scenarios remain stable before and after RALAD fine-tuning, while in simulated environments,the mIOU and mAP metrics are improved by 10.30% and 12.29%, respectively. Moreover, the re-training cost of our approach is reduced by approximately 88.1%. Our code is available at https://github.com/JiachengZuo/RALAD.git.

View Paper

Can We Generate Images with CoT? Let's Verify and Reinforce Image Generation Step by Step

Authors: Ziyu Guo, Renrui Zhang, Chengzhuo Tong, Zhizheng Zhao, Rui Huang, Haoquan Zhang, Manyuan Zhang, Jiaming Liu, Shanghang Zhang, Peng Gao, Hongsheng Li, Pheng-Ann Heng

Abstract: arXiv:2501.13926v2 Announce Type: replace-cross Abstract: Chain-of-Thought (CoT) reasoning has been extensively explored in large models to tackle complex understanding tasks. However, it still remains an open question whether such strategies can be applied to verifying and reinforcing image generation scenarios. In this paper, we provide the first comprehensive investigation of the potential of CoT reasoning to enhance autoregressive image generation. We focus on three techniques: scaling test-time computation for verification, aligning model preferences with Direct Preference Optimization (DPO), and integrating these techniques for complementary effects. Our results demonstrate that these approaches can be effectively adapted and combined to significantly improve image generation performance. Furthermore, given the pivotal role of reward models in our findings, we propose the Potential Assessment Reward Model (PARM) and PARM++, specialized for autoregressive image generation. PARM adaptively assesses each generation step through a potential assessment approach, merging the strengths of existing reward models, and PARM++ further introduces a reflection mechanism to self-correct the generated unsatisfactory image, which is the first to incorporate reflection in autoregressive image generation. Using our investigated reasoning strategies, we enhance a baseline model, Show-o, to achieve superior results, with a significant +24% improvement on the GenEval benchmark, surpassing Stable Diffusion 3 by +15%. We hope our study provides unique insights and paves a new path for integrating CoT reasoning with autoregressive image generation. Code and models are released at https://github.com/ZiyuGuo99/Image-Generation-CoT

View Paper

Optimizing Privacy-Utility Trade-off in Decentralized Learning with Generalized Correlated Noise

Authors: Angelo Rodio, Zheng Chen, Erik G. Larsson

Abstract: arXiv:2501.14644v2 Announce Type: replace-cross Abstract: Decentralized learning enables distributed agents to collaboratively train a shared machine learning model without a central server, through local computation and peer-to-peer communication. Although each agent retains its dataset locally, sharing local models can still expose private information about the local training datasets to adversaries. To mitigate privacy attacks, a common strategy is to inject random artificial noise at each agent before exchanging local models between neighbors. However, this often leads to utility degradation due to the negative effects of cumulated artificial noise on the learning algorithm. In this work, we introduce CorN-DSGD, a novel covariance-based framework for generating correlated privacy noise across agents, which unifies several state-of-the-art methods as special cases. By leveraging network topology and mixing weights, CorN-DSGD optimizes the noise covariance to achieve network-wide noise cancellation. Experimental results show that CorN-DSGD cancels more noise than existing pairwise correlation schemes, improving model performance under formal privacy guarantees.

View Paper

Graph Neural Networks for O-RAN Mobility Management: A Link Prediction Approach

Authors: Ana Gonzalez Bermudez, Miquel Farreras, Milan Groshev, Jos\'e Antonio Trujillo, Isabel de la Bandera, Raquel Barco

Abstract: arXiv:2502.02170v2 Announce Type: replace-cross Abstract: Mobility performance has been a key focus in cellular networks up to 5G. To enhance handover (HO) performance, 3GPP introduced Conditional Handover (CHO) and Layer 1/Layer 2 Triggered Mobility (LTM) mechanisms in 5G. While these reactive HO strategies address the trade-off between HO failures (HOF) and ping-pong effects, they often result in inefficient radio resource utilization due to additional HO preparations. To overcome these challenges, this article proposes a proactive HO framework for mobility management in O-RAN, leveraging user-cell link predictions to identify the optimal target cell for HO. We explore various categories of Graph Neural Networks (GNNs) for link prediction and analyze the complexity of applying them to the mobility management domain. Two GNN models are compared using a real-world dataset, with experimental results demonstrating their ability to capture the dynamic and graph-structured nature of cellular networks. Finally, we present key insights from our study and outline future steps to enable the integration of GNN-based link prediction for mobility management in O-RAN networks.

View Paper

RAPID-Net: Accurate Pocket Identification for Binding-Site-Agnostic Docking

Authors: Yaroslav Balytskyi, Inna Hubenko, Alina Balytska, Christopher V. Kelly

Abstract: arXiv:2502.02371v2 Announce Type: replace-cross Abstract: Accurate identification of druggable pockets and their features is essential for structure-based drug design and effective downstream docking. Here, we present RAPID-Net, a deep learning-based algorithm designed for the accurate prediction of binding pockets and seamless integration with docking pipelines. On the PoseBusters benchmark, RAPID-Net-guided AutoDock Vina achieves 54.9% of Top-1 poses with RMSD < 2 A and satisfying the PoseBusters chemical-validity criterion, compared to 49.1% for DiffBindFR. On the most challenging time split of PoseBusters aiming to assess generalization ability (structures submitted after September 30, 2021), RAPID-Net-guided AutoDock Vina achieves 53.1% of Top-1 poses with RMSD < 2 A and PB-valid, versus 59.5% for AlphaFold 3. Notably, in 92.2% of cases, RAPID-Net-guided Vina samples at least one pose with RMSD < 2 A (regardless of its rank), indicating that pose ranking, rather than sampling, is the primary accuracy bottleneck. The lightweight inference, scalability, and competitive accuracy of RAPID-Net position it as a viable option for large-scale virtual screening campaigns. Across diverse benchmark datasets, RAPID-Net outperforms other pocket prediction tools, including PUResNet and Kalasanty, in both docking accuracy and pocket-ligand intersection rates. Furthermore, we demonstrate the potential of RAPID-Net to accelerate the development of novel therapeutics by highlighting its performance on pharmacologically relevant targets. RAPID-Net accurately identifies distal functional sites, offering new opportunities for allosteric inhibitor design. In the case of the RNA-dependent RNA polymerase of SARS-CoV-2, RAPID-Net uncovers a wider array of potential binding pockets than existing predictors, which typically annotate only the orthosteric pocket and overlook secondary cavities.

View Paper

Koel-TTS: Enhancing LLM based Speech Generation with Preference Alignment and Classifier Free Guidance

Authors: Shehzeen Hussain, Paarth Neekhara, Xuesong Yang, Edresson Casanova, Subhankar Ghosh, Mikyas T. Desta, Roy Fejgin, Rafael Valle, Jason Li

Abstract: arXiv:2502.05236v2 Announce Type: replace-cross Abstract: While autoregressive speech token generation models produce speech with remarkable variety and naturalness, their inherent lack of controllability often results in issues such as hallucinations and undesired vocalizations that do not conform to conditioning inputs. We introduce Koel-TTS, a suite of enhanced encoder-decoder Transformer TTS models that address these challenges by incorporating preference alignment techniques guided by automatic speech recognition and speaker verification models. Additionally, we incorporate classifier-free guidance to further improve synthesis adherence to the transcript and reference speaker audio. Our experiments demonstrate that these optimizations significantly enhance target speaker similarity, intelligibility, and naturalness of synthesized speech. Notably, Koel-TTS directly maps text and context audio to acoustic tokens, and on the aforementioned metrics, outperforms state-of-the-art TTS models, despite being trained on a significantly smaller dataset. Audio samples and demos are available on our website.

View Paper

An Efficient and Precise Training Data Construction Framework for Process-supervised Reward Model in Mathematical Reasoning

Authors: Wei Sun, Qianlong Du, Fuwei Cui, Jiajun Zhang

Abstract: arXiv:2503.02382v2 Announce Type: replace-cross Abstract: Enhancing the mathematical reasoning capabilities of Large Language Models (LLMs) is of great scientific and practical significance. Researchers typically employ process-supervised reward models (PRMs) to guide the reasoning process, effectively improving the models' reasoning abilities. However, existing methods for constructing process supervision training data, such as manual annotation and per-step Monte Carlo estimation, are often costly or suffer from poor quality. To address these challenges, this paper introduces a framework called EpicPRM, which annotates each intermediate reasoning step based on its quantified contribution and uses an adaptive binary search algorithm to enhance both annotation precision and efficiency. Using this approach, we efficiently construct a high-quality process supervision training dataset named Epic50k, consisting of 50k annotated intermediate steps. Compared to other publicly available datasets, the PRM trained on Epic50k demonstrates significantly superior performance. Getting Epic50k at https://github.com/xiaolizh1/EpicPRM.

View Paper

AlignDistil: Token-Level Language Model Alignment as Adaptive Policy Distillation

Authors: Songming Zhang, Xue Zhang, Tong Zhang, Bojie Hu, Yufeng Chen, Jinan Xu

Abstract: arXiv:2503.02832v3 Announce Type: replace-cross Abstract: In modern large language models (LLMs), LLM alignment is of crucial importance and is typically achieved through methods such as reinforcement learning from human feedback (RLHF) and direct preference optimization (DPO). However, in most existing methods for LLM alignment, all tokens in the response are optimized using a sparse, response-level reward or preference annotation. The ignorance of token-level rewards may erroneously punish high-quality tokens or encourage low-quality tokens, resulting in suboptimal performance and slow convergence speed. To address this issue, we propose AlignDistil, an RLHF-equivalent distillation method for token-level reward optimization. Specifically, we introduce the reward learned by DPO into the RLHF objective and theoretically prove the equivalence between this objective and a token-level distillation process, where the teacher distribution linearly combines the logits from the DPO model and a reference model. On this basis, we further bridge the accuracy gap between the reward from the DPO model and the pure reward model, by building a contrastive DPO reward with a normal and a reverse DPO model. Moreover, to avoid under- and over-optimization on different tokens, we design a token adaptive logit extrapolation mechanism to construct an appropriate teacher distribution for each token. Experimental results demonstrate the superiority of our AlignDistil over existing methods and showcase fast convergence due to its token-level distributional reward optimization.

View Paper

ORANSight-2.0: Foundational LLMs for O-RAN

Authors: Pranshav Gajjar, Vijay K. Shah

Abstract: arXiv:2503.05200v2 Announce Type: replace-cross Abstract: Despite the transformative impact of Large Language Models (LLMs) across critical domains such as healthcare, customer service, and business marketing, their integration into Open Radio Access Networks (O-RAN) remains limited. This gap is primarily due to the absence of domain-specific foundational models, with existing solutions often relying on general-purpose LLMs that fail to address the unique challenges and technical intricacies of O-RAN. To bridge this gap, we introduce ORANSight-2.0 (O-RAN Insights), a pioneering initiative to develop specialized foundational LLMs tailored for O-RAN. Built on 18 models spanning five open-source LLM frameworks -- Mistral, Qwen, Llama, Phi, and Gemma -- ORANSight-2.0 fine-tunes models ranging from 1B to 70B parameters, significantly reducing reliance on proprietary, closed-source models while enhancing performance in O-RAN-specific tasks. At the core of ORANSight-2.0 is RANSTRUCT, a novel Retrieval-Augmented Generation (RAG)-based instruction-tuning framework that employs two LLM agents -- a Mistral-based Question Generator and a Qwen-based Answer Generator -- to create high-quality instruction-tuning datasets. The generated dataset is then used to fine-tune the 18 pre-trained open-source LLMs via QLoRA. To evaluate ORANSight-2.0, we introduce srsRANBench, a novel benchmark designed for code generation and codebase understanding in the context of srsRAN, a widely used 5G O-RAN stack.

View Paper

A Deep Learning Approach for Augmenting Perceptional Understanding of Histopathology Images

Authors: Xiaoqian Hu

Abstract: arXiv:2503.06894v3 Announce Type: replace-cross Abstract: In Recent Years, Digital Technologies Have Made Significant Strides In Augmenting-Human-Health, Cognition, And Perception, Particularly Within The Field Of Computational-Pathology. This Paper Presents A Novel Approach To Enhancing The Analysis Of Histopathology Images By Leveraging A Mult-modal-Model That Combines Vision Transformers (Vit) With Gpt-2 For Image Captioning. The Model Is Fine-Tuned On The Specialized Arch-Dataset, Which Includes Dense Image Captions Derived From Clinical And Academic Resources, To Capture The Complexities Of Pathology Images Such As Tissue Morphologies, Staining Variations, And Pathological Conditions. By Generating Accurate, Contextually Captions, The Model Augments The Cognitive Capabilities Of Healthcare Professionals, Enabling More Efficient Disease Classification, Segmentation, And Detection. The Model Enhances The Perception Of Subtle Pathological Features In Images That Might Otherwise Go Unnoticed, Thereby Improving Diagnostic Accuracy. Our Approach Demonstrates The Potential For Digital Technologies To Augment Human Cognitive Abilities In Medical Image Analysis, Providing Steps Toward More Personalized And Accurate Healthcare Outcomes.

View Paper

Att-Adapter: A Robust and Precise Domain-Specific Multi-Attributes T2I Diffusion Adapter via Conditional Variational Autoencoder

Authors: Wonwoong Cho, Yan-Ying Chen, Matthew Klenk, David I. Inouye, Yanxia Zhang

Abstract: arXiv:2503.11937v3 Announce Type: replace-cross Abstract: Text-to-Image (T2I) Diffusion Models have achieved remarkable performance in generating high quality images. However, enabling precise control of continuous attributes, especially multiple attributes simultaneously, in a new domain (e.g., numeric values like eye openness or car width) with text-only guidance remains a significant challenge. To address this, we introduce the Attribute (Att) Adapter, a novel plug-and-play module designed to enable fine-grained, multi-attributes control in pretrained diffusion models. Our approach learns a single control adapter from a set of sample images that can be unpaired and contain multiple visual attributes. The Att-Adapter leverages the decoupled cross attention module to naturally harmonize the multiple domain attributes with text conditioning. We further introduce Conditional Variational Autoencoder (CVAE) to the Att-Adapter to mitigate overfitting, matching the diverse nature of the visual world. Evaluations on two public datasets show that Att-Adapter outperforms all LoRA-based baselines in controlling continuous attributes. Additionally, our method enables a broader control range and also improves disentanglement across multiple attributes, surpassing StyleGAN-based techniques. Notably, Att-Adapter is flexible, requiring no paired synthetic data for training, and is easily scalable to multiple attributes within a single model.

View Paper

ICCO: Learning an Instruction-conditioned Coordinator for Language-guided Task-aligned Multi-robot Control

Authors: Yoshiki Yano, Kazuki Shibata, Maarten Kokshoorn, Takamitsu Matsubara

Abstract: arXiv:2503.12122v2 Announce Type: replace-cross Abstract: Recent advances in Large Language Models (LLMs) have permitted the development of language-guided multi-robot systems, which allow robots to execute tasks based on natural language instructions. However, achieving effective coordination in distributed multi-agent environments remains challenging due to (1) misalignment between instructions and task requirements and (2) inconsistency in robot behaviors when they independently interpret ambiguous instructions. To address these challenges, we propose Instruction-Conditioned Coordinator (ICCO), a Multi-Agent Reinforcement Learning (MARL) framework designed to enhance coordination in language-guided multi-robot systems. ICCO consists of a Coordinator agent and multiple Local Agents, where the Coordinator generates Task-Aligned and Consistent Instructions (TACI) by integrating language instructions with environmental states, ensuring task alignment and behavioral consistency. The Coordinator and Local Agents are jointly trained to optimize a reward function that balances task efficiency and instruction following. A Consistency Enhancement Term is added to the learning objective to maximize mutual information between instructions and robot behaviors, further improving coordination. Simulation and real-world experiments validate the effectiveness of ICCO in achieving language-guided task-aligned multi-robot control. The demonstration can be found at https://yanoyoshiki.github.io/ICCO/.

View Paper

Towards Detecting Persuasion on Social Media: From Model Development to Insights on Persuasion Strategies

Authors: Elyas Meguellati, Stefano Civelli, Pietro Bernardelle, Shazia Sadiq, Irwin King, Gianluca Demartini

Abstract: arXiv:2503.13844v2 Announce Type: replace-cross Abstract: Political advertising plays a pivotal role in shaping public opinion and influencing electoral outcomes, often through subtle persuasive techniques embedded in broader propaganda strategies. Detecting these persuasive elements is crucial for enhancing voter awareness and ensuring transparency in democratic processes. This paper presents an integrated approach that bridges model development and real-world application through two interconnected studies. First, we introduce a lightweight model for persuasive text detection that achieves state-of-the-art performance in Subtask 3 of SemEval 2023 Task 3 while requiring significantly fewer computational resources and training data than existing methods. Second, we demonstrate the model's practical utility by collecting the Australian Federal Election 2022 Facebook Ads (APA22) dataset, partially annotating a subset for persuasion, and fine-tuning the model to adapt from mainstream news to social media content. We then apply the fine-tuned model to label the remainder of the APA22 dataset, revealing distinct patterns in how political campaigns leverage persuasion through different funding strategies, word choices, demographic targeting, and temporal shifts in persuasion intensity as election day approaches.

Our findings not only underscore the necessity of domain-specific modeling for analyzing persuasion on social media but also show how uncovering these strategies can enhance transparency, inform voters, and promote accountability in digital campaigns.

View Paper

AirCache: Activating Inter-modal Relevancy KV Cache Compression for Efficient Large Vision-Language Model Inference

Authors: Kai Huang, Hao Zou, Bochen Wang, Ye Xi, Zhen Xie, Hao Wang

Abstract: arXiv:2503.23956v3 Announce Type: replace-cross Abstract: Recent advancements in Large Visual Language Models (LVLMs) have gained significant attention due to their remarkable reasoning capabilities and proficiency in generalization. However, processing a large number of visual tokens and generating long-context outputs impose substantial computational overhead, leading to excessive demands for key-value (KV) cache. To address this critical bottleneck, we propose AirCache, a novel KV cache compression method aimed at accelerating LVLMs inference. This work systematically investigates the correlations between visual and textual tokens within the attention mechanisms of LVLMs. Our empirical analysis reveals considerable redundancy in cached visual tokens, wherein strategically eliminating these tokens preserves model performance while significantly accelerating context generation. Inspired by these findings, we introduce an elite observation window for assessing the importance of visual components in the KV cache, focusing on stable inter-modal relevancy modeling with enhanced multi-perspective consistency. Additionally, we develop an adaptive layer-wise budget allocation strategy that capitalizes on the strength and skewness of token importance distribution, showcasing superior efficiency compared to uniform allocation. Comprehensive evaluations across multiple LVLMs and benchmarks demonstrate that our method achieves comparable performance to the full cache while retaining only 10% of visual KV cache, thereby reducing decoding latency by 29% to 66% across various batch size and prompt length of inputs. Notably, as cache retention rates decrease, our method exhibits increasing performance advantages over existing approaches.

View Paper

Attention-Based Multiscale Temporal Fusion Network for Uncertain-Mode Fault Diagnosis in Multimode Processes

Authors: Guangqiang Li, M. Amine Atoui, Xiangshun Li

Abstract: arXiv:2504.05172v3 Announce Type: replace-cross Abstract: Fault diagnosis in multimode processes plays a critical role in ensuring the safe operation of industrial systems across multiple modes. It faces a great challenge yet to be addressed - that is, the significant distributional differences among monitoring data from multiple modes make it difficult for the models to extract shared feature representations related to system health conditions. In response to this problem, this paper introduces a novel method called attention-based multiscale temporal fusion network. The multiscale depthwise convolution and gated recurrent unit are employed to extract multiscale contextual local features and long-short-term features. Instance normalization is applied to suppress mode-specific information. Furthermore, a temporal attention mechanism is designed to focus on critical time points with higher cross-mode shared information, thereby enhancing the accuracy of fault diagnosis. The proposed model is applied to Tennessee Eastman process dataset and three-phase flow facility dataset. The experiments demonstrate that the proposed model achieves superior diagnostic performance and maintains a small model size. The source code will be available on GitHub at https://github.com/GuangqiangLi/AMTFNet.

View Paper

Parasite: A Steganography-based Backdoor Attack Framework for Diffusion Models

Authors: Jiahao Chen, Yu Pan, Yi Du, Chunkai Wu, Lin Wang

Abstract: arXiv:2504.05815v2 Announce Type: replace-cross Abstract: Recently, the diffusion model has gained significant attention as one of the most successful image generation models, which can generate high-quality images by iteratively sampling noise. However, recent studies have shown that diffusion models are vulnerable to backdoor attacks, allowing attackers to enter input data containing triggers to activate the backdoor and generate their desired output. Existing backdoor attack methods primarily focused on target noise-to-image and text-to-image tasks, with limited work on backdoor attacks in image-to-image tasks. Furthermore, traditional backdoor attacks often rely on a single, conspicuous trigger to generate a fixed target image, lacking concealability and flexibility. To address these limitations, we propose a novel backdoor attack method called "Parasite" for image-to-image tasks in diffusion models, which not only is the first to leverage steganography for triggers hiding, but also allows attackers to embed the target content as a backdoor trigger to achieve a more flexible attack. "Parasite" as a novel attack method effectively bypasses existing detection frameworks to execute backdoor attacks. In our experiments, "Parasite" achieved a 0 percent backdoor detection rate against the mainstream defense frameworks. In addition, in the ablation study, we discuss the influence of different hiding coefficients on the attack results. You can find our code at https://anonymous.4open.science/r/Parasite-1715/.

View Paper

Mapping Industry Practices to the EU AI Act's GPAI Code of Practice Safety and Security Measures

Authors: Lily Stelling, Mick Yang, Rokas Gipi\v{s}kis, Leon Staufer, Ze Shen Chin, Sim\'eon Campos, Ariel Gil, Michael Chen

Abstract: arXiv:2504.15181v2 Announce Type: replace-cross Abstract: This report provides a detailed comparison between the Safety and Security measures proposed in the EU AI Act's General-Purpose AI (GPAI) Code of Practice (Third Draft) and the current commitments and practices voluntarily adopted by leading AI companies. As the EU moves toward enforcing binding obligations for GPAI model providers, the Code of Practice will be key for bridging legal requirements with concrete technical commitments. Our analysis focuses on the draft's Safety and Security section (Commitments II.1-II.16), documenting excerpts from current public-facing documents that are relevant to each individual measure. We systematically reviewed different document types, such as companies' frontier safety frameworks and model cards, from over a dozen companies, including OpenAI, Anthropic, Google DeepMind, Microsoft, Meta, Amazon, and others. This report is not meant to be an indication of legal compliance, nor does it take any prescriptive viewpoint about the Code of Practice or companies' policies. Instead, it aims to inform the ongoing dialogue between regulators and General-Purpose AI model providers by surfacing evidence of industry precedent for various measures. Nonetheless, we were able to find relevant quotes from at least 5 companies' documents for the majority of the measures in Commitments II.1-II.16.

View Paper

DMS-Net:Dual-Modal Multi-Scale Siamese Network for Binocular Fundus Image Classification

Authors: Guohao Huo, Zibo Lin, Zitong Wang, Ruiting Dai, Hao Tang

Abstract: arXiv:2504.18046v2 Announce Type: replace-cross Abstract: Ophthalmic diseases pose a significant global health challenge, yet traditional diagnosis methods and existing single-eye deep learning approaches often fail to account for binocular pathological correlations. To address this, we propose DMS-Net, a dual-modal multi-scale Siamese network for binocular fundus image classification. Our framework leverages weight-shared Siamese ResNet-152 backbones to extract deep semantic features from paired fundus images. To tackle challenges such as lesion boundary ambiguity and scattered pathological distributions, we introduce a Multi-Scale Context-Aware Module (MSCAM) that integrates adaptive pooling and attention mechanisms for multi-resolution feature aggregation. Additionally, a Dual-Modal Feature Fusion (DMFF) module enhances cross-modal interaction through spatial-semantic recalibration and bidirectional attention, effectively combining global context and local edge features. Evaluated on the ODIR-5K dataset, DMS-Net achieves state-of-the-art performance with 82.9% accuracy, 84.5% recall, and 83.2% Cohen's kappa, demonstrating superior capability in detecting symmetric pathologies and advancing clinical decision-making for ocular diseases.

View Paper

Machine Learning-Based Modeling of the Anode Heel Effect in X-ray Beam Monte Carlo Simulations

Authors: Hussein Harb, Didier Benoit, Axel Rannou, Chi-Hieu Pham, Valentin Tissot, Bahaa Nasr, Julien Bert

Abstract: arXiv:2504.19155v2 Announce Type: replace-cross Abstract: To develop a machine learning-based framework for accurately modeling the anode heel effect in Monte Carlo simulations of X-ray imaging systems, enabling realistic beam intensity profiles with minimal experimental calibration. Multiple regression models were trained to predict spatial intensity variations along the anode-cathode axis using experimentally acquired weights derived from beam measurements across different tube potentials. These weights captured the asymmetry introduced by the anode heel effect. A systematic fine-tuning protocol was established to minimize the number of required measurements while preserving model accuracy. The models were implemented in the OpenGATE 10 and GGEMS Monte Carlo toolkits to evaluate their integration feasibility and predictive performance. Among the tested models, gradient boosting regression (GBR) delivered the highest accuracy, with prediction errors remaining below 5% across all energy levels. The optimized fine-tuning strategy required only six detector positions per energy level, reducing measurement effort by 65%. The maximum error introduced through this fine-tuning process remained below 2%. Dose actor comparisons within Monte Carlo simulations demonstrated that the GBR-based model closely replicated clinical beam profiles and significantly

outperformed conventional symmetric beam models. This study presents a robust and generalizable method for incorporating the anode heel effect into Monte Carlo simulations using machine learning. By enabling accurate, energy-dependent beam modeling with limited calibration data, the approach enhances simulation realism for applications in clinical dosimetry, image quality assessment, and radiation protection.

View Paper

Monitoring digestate application on agricultural crops using Sentinel-2 Satellite imagery

Authors: Andreas Kalogeras, Dimitrios Bormpoudakis, Iason Tsardanidis, Dimitra A. Loka, Charalampos Kontoes

Abstract: arXiv:2504.19996v2 Announce Type: replace-cross Abstract: The widespread use of Exogenous Organic Matter in agriculture necessitates monitoring to assess its effects on soil and crop health. This study evaluates optical Sentinel-2 satellite imagery for detecting digestate application, a practice that enhances soil fertility but poses environmental risks like microplastic contamination and nitrogen losses. In the first instance, Sentinel-2 satellite image time series (SITS) analysis of specific indices (EOMI, NDVI, EVI) was used to characterize EOM's spectral behavior after application on the soils of four different crop types in Thessaly, Greece. Furthermore, Machine Learning (ML) models (namely Random Forest, k-NN, Gradient Boosting and a Feed-Forward Neural Network), were used to investigate digestate presence detection, achieving F1-scores up to 0.85. The findings highlight the potential of combining remote sensing and ML for scalable and cost-effective monitoring of EOM applications, supporting precision agriculture and sustainability.

View Paper

RoBridge: A Hierarchical Architecture Bridging Cognition and Execution for General Robotic Manipulation

Authors: Kaidong Zhang, Rongtao Xu, Pengzhen Ren, Junfan Lin, Hefeng Wu, Liang Lin, Xiaodan Liang

Abstract: arXiv:2505.01709v3 Announce Type: replace-cross Abstract: Operating robots in open-ended scenarios with diverse tasks is a crucial research and application direction in robotics. While recent progress in natural language processing and large multimodal models has enhanced robots' ability to

understand complex instructions, robot manipulation still faces the procedural skill dilemma and the declarative skill dilemma in open environments. Existing methods often compromise cognitive and executive capabilities. To address these challenges, in this paper, we propose RoBridge, a hierarchical intelligent architecture for general robotic manipulation. It consists of a high-level cognitive planner (HCP) based on a large-scale pre-trained vision-language model (VLM), an invariant operable representation (IOR) serving as a symbolic bridge, and a generalist embodied agent (GEA). RoBridge maintains the declarative skill of VLM and unleashes the procedural skill of reinforcement learning, effectively bridging the gap between cognition and execution. RoBridge demonstrates significant performance improvements over existing baselines, achieving a 75% success rate on new tasks and an 83% average success rate in sim-to-real generalization using only five real-world data samples per task. This work represents a significant step towards integrating cognitive reasoning with physical execution in robotic systems, offering a new paradigm for general robotic manipulation.

View Paper

Application of YOLOv8 in monocular downward multiple Car Target detection

Authors: Shijie Lyu

Abstract: arXiv:2505.10016v2 Announce Type: replace-cross Abstract: Autonomous driving technology is progressively transforming traditional car driving methods, marking a significant milestone in modern transportation. Object detection serves as a cornerstone of autonomous systems, playing a vital role in enhancing driving safety, enabling autonomous functionality, improving traffic efficiency, and facilitating effective emergency responses. However, current technologies such as radar for environmental perception, cameras for road perception, and vehicle sensor networks face notable challenges, including high costs, vulnerability to weather and lighting conditions, and limited resolution.To address these limitations, this paper presents an improved autonomous target detection network based on YOLOv8. By integrating structural reparameterization technology, a bidirectional pyramid structure network model, and a novel detection pipeline into the YOLOv8 framework, the proposed approach achieves highly efficient and precise detection of multi-scale, small, and remote objects. Experimental results demonstrate that the enhanced model can effectively detect both large and small objects with a detection accuracy of 65%, showcasing significant advancements over traditional methods.This improved model holds substantial potential for real-world applications and is well-suited for autonomous driving competitions, such as the Formula Student Autonomous China (FSAC), particularly excelling in scenarios involving single-target and small-object detection.

View Paper

ORL-LDM: Offline Reinforcement Learning Guided Latent Diffusion Model Super-Resolution Reconstruction

Authors: Shijie Lyu

Abstract: arXiv:2505.10027v2 Announce Type: replace-cross Abstract: With the rapid advancement of remote sensing technology, super-resolution image reconstruction is of great research and practical significance. Existing deep learning methods have made progress but still face limitations in handling complex scenes and preserving image details. This paper proposes a reinforcement learning-based latent diffusion model (LDM) fine-tuning method for remote sensing image super-resolution. The method constructs a reinforcement learning environment with states, actions, and rewards, optimizing decision objectives through proximal policy optimization (PPO) during the reverse denoising process of the LDM model. Experiments on the RESISC45 dataset show significant improvements over the baseline model in PSNR, SSIM, and LPIPS, with PSNR increasing by 3-4dB, SSIM improving by 0.08-0.11, and LPIPS reducing by 0.06-0.10, particularly in structured and complex natural scenes. The results demonstrate the method's effectiveness in enhancing super-resolution quality and adaptability across scenes.

View Paper

Deep Video Discovery: Agentic Search with Tool Use for Long-form Video Understanding

Authors: Xiaoyi Zhang, Zhaoyang Jia, Zongyu Guo, Jiahao Li, Bin Li, Houqiang Li, Yan Lu

Abstract: arXiv:2505.18079v3 Announce Type: replace-cross Abstract: Long-form video understanding presents significant challenges due to extensive temporal-spatial complexity and the difficulty of question answering under such extended contexts. While Large Language Models (LLMs) have demonstrated considerable advancements in video analysis capabilities and long context handling, they continue to exhibit limitations when processing information-dense hour-long videos. To overcome such limitations, we propose the Deep Video Discovery agent to leverage an agentic search strategy over segmented video clips. Different from previous video agents manually designing a rigid workflow, our approach emphasizes the autonomous nature of agents. By providing a set of search-centric tools on multi-granular video database, our DVD agent leverages the advanced reasoning capability of LLM to plan on its current observation state, strategically selects tools, formulates appropriate parameters for actions, and iteratively refines its internal reasoning in light of the gathered information. We perform comprehensive evaluation on multiple long video understanding benchmarks that demonstrates the advantage of the entire system design. Our DVD agent achieves SOTA performance, significantly surpassing prior works by a large margin on the challenging LVBench dataset. Comprehensive ablation studies and in-depth tool analyses are also provided, yielding insights to further advance intelligent agents tailored for long-form video understanding tasks. The code has been released in https://github.com/microsoft/DeepVideoDiscovery.

View Paper

Robot Operation of Home Appliances by Reading User Manuals

Authors: Jian Zhang, Hanbo Zhang, Anxing Xiao, David Hsu

Abstract: arXiv:2505.20424v2 Announce Type: replace-cross Abstract: Operating home appliances, among the most common tools in every household, is a critical capability for assistive home robots. This paper presents ApBot, a robot system that operates novel household appliances by "reading" their user manuals. ApBot faces multiple challenges: (i) infer goal-conditioned partial policies from their unstructured, textual descriptions in a user manual document, (ii) ground the policies to the appliance in the physical world, and (iii) execute the policies reliably over potentially many steps, despite compounding errors. To tackle these challenges, ApBot constructs a structured, symbolic model of an appliance from its manual, with the help of a large vision-language model (VLM). It grounds the symbolic actions visually to control panel elements. Finally, ApBot closes the loop by updating the model based on visual feedback. Our experiments show that across a wide range of simulated and real-world appliances, ApBot achieves consistent and statistically significant improvements in task success rate, compared with state-of-the-art large VLMs used directly as control policies. These results suggest that a structured internal representations plays an important role in robust robot operation of home appliances, especially, complex ones.

View Paper

Advancing Multimodal Reasoning via Reinforcement Learning with Cold Start

Authors: Lai Wei, Yuting Li, Kaipeng Zheng, Chen Wang, Yue Wang, Linghe Kong, Lichao Sun, Weiran Huang

Abstract: arXiv:2505.22334v2 Announce Type: replace-cross Abstract: Recent advancements in large language models (LLMs) have demonstrated impressive chain-of-thought reasoning capabilities, with reinforcement learning (RL) playing a crucial role in this progress. While "aha moment" patterns--where models exhibit self-correction through reflection--are often attributed to emergent properties from RL, we first demonstrate that these patterns exist in multimodal LLMs (MLLMs) prior to RL training but may not necessarily correlate with improved reasoning performance. Building on these insights, we present a comprehensive study on enhancing multimodal reasoning through a two-stage approach: (1) supervised fine-tuning (SFT) as a cold start with structured chain-of-thought reasoning patterns, followed by (2) reinforcement learning via GRPO to further refine these capabilities. Our extensive experiments show that this combined approach consistently outperforms both SFT-only and RL-only methods across challenging multimodal reasoning benchmarks. The resulting models achieve state-of-the-art performance among open-source MLLMs at both 3B and 7B scales, with our 7B model showing substantial improvements over base models (e.g., 66.3 %$\rightarrow$73.4 % on MathVista, 62.9 %$ \rightarrow$70.4 % on We-Math) and our 3B model achieving performance competitive with several 7B models. Overall, this work provides practical guidance for building advanced multimodal reasoning models. Our code is available at https://github.com/waltonfuture/RL-with-Cold-Start.

View Paper

Unified Sparse-Matrix Representations for Diverse Neural Architectures

Authors: Yuzhou Zhu

Abstract: arXiv:2506.01966v3 Announce Type: replace-cross Abstract: Deep neural networks employ specialized architectures for vision, sequential and language tasks, yet this proliferation obscures their underlying commonalities. We introduce a unified matrix-order framework that casts convolutional, recurrent and self-attention operations as sparse matrix multiplications. Convolution is realized via an upper-triangular weight matrix performing first-order transformations; recurrence emerges from a lower-triangular matrix encoding stepwise updates; attention arises naturally as a third-order tensor factorization. We prove algebraic isomorphism with standard CNN, RNN and Transformer layers under mild assumptions. Empirical evaluations on image classification (MNIST, CIFAR-10/100, Tiny ImageNet), time-series forecasting (ETTh1, Electricity Load Diagrams) and language modeling/classification (AG News, WikiText-2, Penn Treebank) confirm that sparse-matrix formulations match or exceed native model performance while converging in comparable or fewer epochs. By reducing architecture design to sparse pattern selection, our matrix perspective aligns with GPU parallelism and leverages mature algebraic optimization tools. This work establishes a mathematically rigorous substrate for diverse neural architectures and opens avenues for principled, hardware-aware network design.

View Paper

KVCache Cache in the Wild: Characterizing and Optimizing KVCache Cache at a Large Cloud Provider

Authors: Jiahao Wang, Jinbo Han, Xingda Wei, Sijie Shen, Dingyan Zhang, Chenguang Fang, Rong Chen, Wenyuan Yu, Haibo Chen

Abstract: arXiv:2506.02634v4 Announce Type: replace-cross Abstract: Serving large language models (LLMs) is important for cloud providers, and caching intermediate results (KV\$) after processing each request substantially improves serving throughput and latency. However, there is limited understanding of how LLM serving benefits from KV\$ caching, where system design decisions like cache eviction policies are highly workload-dependent. In this paper, we present the first systematic characterization of the KV\$ workload patterns from one of the leading LLM service providers. We draw observations that were not covered by previous studies focusing on synthetic workloads, including: KV\$ reuses are skewed across requests, where reuses between single-turn requests are equally important as multi-turn requests; the reuse time and probability are diverse considering all requests, but for a specific request category, the pattern tends to be predictable; and the overall cache size required for an ideal cache hit ratio is moderate. Based on the characterization, we further propose a workload-aware cache eviction policy that improves the serving performance under real-world traces, especially with limited cache capacity.

View Paper

MRI-CORE: A Foundation Model for Magnetic Resonance Imaging

Authors: Haoyu Dong, Yuwen Chen, Hanxue Gu, Nicholas Konz, Yaqian Chen, Qihang Li, Maciej A. Mazurowski

Abstract: arXiv:2506.12186v2 Announce Type: replace-cross Abstract: The widespread use of Magnetic Resonance Imaging (MRI) in combination with deep learning shows promise for many high-impact automated diagnostic and prognostic tools. However, training new models requires large amounts of labeled data, a challenge due to high cost of precise annotations and data privacy. To address this issue, we introduce the MRI-CORE, a vision foundation model trained using more than 6 million slices from over 110 thousand MRI volumes across 18 body locations. Our experiments show notable improvements in performance over state-of-the-art methods in 13 data-restricted segmentation tasks, as well as in image classification, and zero-shot segmentation, showing the strong potential of MRI-CORE to enable data-efficient development of artificial intelligence models. We also present data on which strategies yield most useful foundation models and a novel analysis relating similarity between pre-training and downstream task data with transfer learning performance. Our model is publicly available with a permissive license.

View Paper

EXGnet: a single-lead explainable-AI guided multiresolution network with train-only quantitative features for trustworthy ECG arrhythmia classification

Authors: Tushar Talukder Showrav, Soyabul Islam Lincoln, Md. Kamrul Hasan

Abstract: arXiv:2506.12404v2 Announce Type: replace-cross Abstract: Deep learning has significantly propelled the performance of ECG arrhythmia classification, yet its clinical adoption remains hindered by challenges in interpretability and deployment on resource-constrained edge devices. To bridge this gap, we propose EXGnet, a novel and reliable ECG arrhythmia classification network tailored for single-lead signals, specifically designed to balance high accuracy, explainability, and edge compatibility. EXGnet integrates XAI supervision during training via a normalized cross-correlation based loss, directing the model's attention to clinically relevant ECG regions, similar to a cardiologist's focus. This supervision is driven by automatically generated ground truth, derived through an innovative heart rate variability-based approach, without the need for manual annotation. To enhance classification accuracy without compromising deployment simplicity, we incorporate quantitative ECG features during training. These enrich the model with multi-domain knowledge but are excluded during inference, keeping the model lightweight for edge deployment. Additionally, we introduce an innovative multiresolution block to efficiently capture both short and long-term signal features while maintaining computational efficiency. Rigorous evaluation on the Chapman and Ningbo benchmark datasets validates the supremacy of EXGnet, which achieves average five-fold accuracies of 98.762% and 96.932%, and F1-scores of 97.910% and 95.527%, respectively. Comprehensive ablation studies and both quantitative and qualitative interpretability assessment confirm that the XAI guidance is pivotal, demonstrably enhancing the model's focus and trustworthiness. Overall, EXGnet sets a new benchmark by combining high-performance arrhythmia classification with interpretability, paving the way for more trustworthy and accessible portable ECG based health monitoring systems.

View Paper

LoX: Low-Rank Extrapolation Robustifies LLM Safety Against Fine-tuning

Authors: Gabriel J. Perin, Runjin Chen, Xuxi Chen, Nina S. T. Hirata, Zhangyang Wang, Junyuan Hong

Abstract: arXiv:2506.15606v2 Announce Type: replace-cross Abstract: Large Language Models (LLMs) have become indispensable in real-world applications. However, their widespread adoption raises significant safety concerns, particularly in responding to socially harmful questions. Despite substantial efforts to improve model safety through alignment, aligned models can still have their safety protections undermined by subsequent fine-tuning - even when the additional training data appears benign. In this paper, we empirically demonstrate that this vulnerability stems from the sensitivity of safety-critical low-rank subspaces in LLM parameters to fine-tuning. Building on this insight, we propose a novel training-free method, termed Low-Rank Extrapolation (LoX), to enhance safety robustness by extrapolating the safety subspace of an aligned LLM. Our experimental results confirm the effectiveness of LoX, demonstrating significant improvements in robustness against both benign and malicious fine-tuning attacks while preserving the model's adaptability to new tasks. For instance, LoX leads to 11% to 54% absolute reductions in attack success rates (ASR) facing benign or malicious fine-tuning attacks. By investigating the ASR landscape of parameters, we attribute the success of LoX to that the extrapolation moves LLM parameters to a flatter zone, thereby less sensitive to perturbations. The code is available at github.com/VITA-Group/LoX.

View Paper

Modeling Public Perceptions of Science in Media

Authors: Jiaxin Pei, Dustin Wright, Isabelle Augenstein, David Jurgens

Abstract: arXiv:2506.16622v2 Announce Type: replace-cross Abstract: Effectively engaging the public with science is vital for fostering trust and understanding in our scientific community. Yet, with an ever-growing volume of information, science communicators struggle to anticipate how audiences will perceive and interact with scientific news. In this paper, we introduce a computational framework that models public perception across twelve dimensions, such as newsworthiness, importance, and surprisingness. Using this framework, we create a large-scale science news perception dataset with 10,489 annotations from 2,101 participants from diverse US and UK populations, providing valuable insights into public responses to scientific information across domains. We further develop NLP models that predict public perception scores with a strong performance. Leveraging the dataset and model, we examine public perception of science from two perspectives: (1) Perception as an outcome: What factors affect the public perception of scientific information? (2) Perception as a predictor: Can we use the estimated perceptions to predict public engagement with science? We find that individuals' frequency of science news consumption is the driver of perception, whereas demographic factors exert minimal influence. More importantly, through a large-scale analysis and carefully designed natural experiment on Reddit, we demonstrate that the estimated public perception of scientific information has direct connections with the final engagement pattern. Posts with more positive perception scores receive significantly more comments and upvotes, which is consistent across different scientific information and for the same science, but are framed differently. Overall, this research underscores the importance of nuanced perception modeling in science communication, offering new pathways to predict public interest and engagement with scientific content.

View Paper

GTA: Grouped-head latenT Attention

Authors: Luoyang Sun, Cheng Deng, Jiwen Jiang, Xinjian Wu, Haifeng Zhang, Lei Chen, Lionel Ni, Jun Wang

Abstract: arXiv:2506.17286v2 Announce Type: replace-cross Abstract: Attention mechanisms underpin the success of large language models (LLMs), yet their substantial computational and memory overhead poses challenges for optimizing efficiency and performance. A critical bottleneck arises as KV cache and attention computations scale rapidly with text length, challenging deployment on hardware with limited computational and memory resources. We observe that attention mechanisms exhibit substantial redundancy, since the KV cache can be significantly compressed and attention maps across heads display high similarity, revealing that much of the computation and storage is unnecessary. Leveraging these insights, we propose \textbf{G}rouped-Head Laten\textbf{T} \textbf{A}ttention (GTA), a novel attention mechanism that reduces memory usage and computational complexity while maintaining performance. GTA comprises two components: (1) a shared attention map mechanism that reuses attention scores across multiple heads, decreasing the key cache size; and (2) a nonlinear value decoder with learned projections that compresses the value cache into a latent

space, further cutting memory needs. GTA cuts attention computation FLOPs by up to \emph{62.5\%} versus Grouped-Query Attention and shrink the KV cache by up to \emph{70\%}, all while avoiding the extra overhead of Multi-Head Latent Attention to improve LLM deployment efficiency. Consequently, GTA models achieve a \emph{2x} increase in end-to-end inference speed, with prefill benefiting from reduced computational cost and decoding benefiting from the smaller cache footprint.

View Paper

Data-Driven Exploration for a Class of Continuous-Time Indefinite Linear--Quadratic Reinforcement Learning Problems

Authors: Yilie Huang, Xun Yu Zhou

Abstract: arXiv:2507.00358v2 Announce Type: replace-cross Abstract: We study reinforcement learning (RL) for the same class of continuous-time stochastic linear--quadratic (LQ) control problems as in \cite{huang2024sublinear}, where volatilities depend on both states and controls while states are scalar-valued and running control rewards are absent. We propose a model-free, data-driven exploration mechanism that adaptively adjusts entropy regularization by the critic and policy variance by the actor. Unlike the constant or deterministic exploration schedules employed in \cite{huang2024sublinear}, which require extensive tuning for implementations and ignore learning progresses during iterations, our adaptive exploratory approach boosts learning efficiency with minimal tuning. Despite its flexibility, our method achieves a sublinear regret bound that matches the best-known model-free results for this class of LQ problems, which were previously derived only with fixed exploration schedules. Numerical experiments demonstrate that adaptive explorations accelerate convergence and improve regret performance compared to the non-adaptive model-free and model-based counterparts.

View Paper

Prompt Guidance and Human Proximal Perception for HOT Prediction with Regional Joint Loss

Authors: Yuxiao Wang, Yu Lei, Zhenao Wei, Weiying Xue, Xinyu Jiang, Nan Zhuang, Qi Liu

Abstract: arXiv:2507.01630v2 Announce Type: replace-cross Abstract: The task of Human-Object conTact (HOT) detection involves identifying the specific areas of

the human body that are touching objects. Nevertheless, current models are restricted to just one type of image, often leading to too much segmentation in areas with little interaction, and struggling to maintain category consistency within specific regions. To tackle this issue, a HOT framework, termed \textbf{P3HOT}, is proposed, which blends \textbf{P}rompt guidance and human \textbf{P}roximal \textbf{P}erception. To begin with, we utilize a semantic-driven prompt mechanism to direct the network's attention towards the relevant regions based on the correlation between image and text. Then a human proximal perception mechanism is employed to dynamically perceive key depth range around the human, using learnable parameters to effectively eliminate regions where interactions are not expected. Calculating depth resolves the uncertainty of the overlap between humans and objects in a 2D perspective, providing a quasi-3D viewpoint. Moreover, a Regional Joint Loss (RJLoss) has been created as a new loss to inhibit abnormal categories in the same area. A new evaluation metric called ``AD-Acc.'' is introduced to address the shortcomings of existing methods in addressing negative samples. Comprehensive experimental results demonstrate that our approach achieves state-of-the-art performance in four metrics across two benchmark datasets. Specifically, our model achieves an improvement of \textbf{0.7}$\uparrow$, \textbf{2.0}$\uparrow$, \textbf{1.6}$\uparrow$, and \textbf{11.0}$\uparrow$ in SC-Acc., mIoU, wIoU, and AD-Acc. metrics, respectively, on the HOT-Annotated dataset. The sources code are available at https://github.com/YuxiaoWang-AI/P3HOT.

View Paper

SpecCLIP: Aligning and Translating Spectroscopic Measurements for Stars

Authors: Xiaosheng Zhao, Yang Huang, Guirong Xue, Xiao Kong, Jifeng Liu, Xiaoyu Tang, Timothy C. Beers, Yuan-Sen Ting, A-Li Luo

Abstract: arXiv:2507.01939v2 Announce Type: replace-cross Abstract: In recent years, large language models (LLMs) have transformed natural language understanding through vast datasets and large-scale parameterization. Inspired by this success, we present SpecCLIP, a foundation model framework that extends LLM-inspired methodologies to stellar spectral analysis. Stellar spectra, akin to structured language, encode rich physical and chemical information about stars. By training foundation models on large-scale spectral datasets, our goal is to learn robust and informative embeddings that support diverse downstream applications. As a proof of concept, SpecCLIP involves pre-training on two spectral types--LAMOST low-resolution and Gaia XP--followed by contrastive alignment using the CLIP (Contrastive Language-Image Pre-training) framework, adapted to associate spectra from different instruments. This alignment is complemented by auxiliary decoders that preserve spectrum-specific information

and enable translation (prediction) between spectral types, with the former achieved by maximizing mutual information between embeddings and input spectra. The result is a cross-spectrum framework enabling intrinsic calibration and flexible applications across instruments. We demonstrate that fine-tuning these models on moderate-sized labeled datasets improves adaptability to tasks such as stellar-parameter estimation and chemical-abundance determination. SpecCLIP also enhances the accuracy and precision of parameter estimates benchmarked against external survey data. Additionally, its similarity search and cross-spectrum prediction capabilities offer potential for anomaly detection. Our results suggest that contrastively trained foundation models enriched with spectrum-aware decoders can advance precision stellar spectroscopy.

View Paper

How Well Does GPT-4o Understand Vision? Evaluating Multimodal Foundation Models on Standard Computer Vision Tasks

Authors: Rahul Ramachandran, Ali Garjani, Roman Bachmann, Andrei Atanov, O\u{g}uzhan Fatih Kar, Amir Zamir

Abstract: arXiv:2507.01955v2 Announce Type: replace-cross Abstract: Multimodal foundation models, such as GPT-4o, have recently made remarkable progress, but it is not clear where exactly these models stand in terms of understanding vision. In this paper, we benchmark the performance of popular multimodal foundation models (GPT-4o, o4-mini, Gemini 1.5 Pro and Gemini 2.0 Flash, Claude 3.5 Sonnet, Qwen2-VL, Llama 3.2) on standard computer vision tasks (semantic segmentation, object detection, image classification, depth and surface normal prediction) using established datasets (e.g., COCO, ImageNet and its variants, etc). The main challenges to performing this are: 1) most models are trained to output text and cannot natively express versatile domains, such as segments or 3D geometry, and 2) many leading models are proprietary and accessible only at an API level, i.e., there is no weight access to adapt them. We address these challenges by translating standard vision tasks into equivalent text-promptable and API-compatible tasks via prompt chaining to create a standardized benchmarking framework. We observe that 1) the models are not close to the state-of-the-art specialist models at any task. However, 2) they are respectable generalists; this is remarkable as they are presumably trained on primarily image-text-based tasks. 3) They perform semantic tasks notably better than geometric ones. 4) While the prompt-chaining techniques affect performance, better models exhibit less sensitivity to prompt variations. 5) GPT-4o performs the best among non-reasoning models, securing the top position in 4 out of 6 tasks, 6) reasoning models, e.g. o3, show improvements in geometric

tasks, and 7) a preliminary analysis of models with native image generation, like the latest GPT-4o, shows they exhibit quirks like hallucinations and spatial misalignments.

View Paper

Cautious Next Token Prediction

Authors: Yizhou Wang, Lingzhi Zhang, Yue Bai, Mang Tik Chiu, Zhengmian Hu, Mingyuan Zhang, Qihua Dong, Yu Yin, Sohrab Amirghodsi, Yun Fu

Abstract: arXiv:2507.03038v2 Announce Type: replace-cross Abstract: Next token prediction paradigm has been prevailing for autoregressive models in the era of LLMs. The current default sampling choice for popular LLMs is temperature scaling together with nucleus sampling to balance diversity and coherence. Nevertheless, such approach leads to inferior performance in various NLP tasks when the model is not certain about testing questions. To this end, we propose a brand new training-free decoding strategy, dubbed as Cautious Next Token Prediction (CNTP). In the decoding process, if the model has comparatively high prediction entropy at a certain step, we sample multiple trials starting from the step independently and stop when encountering any punctuation. Then we select the trial with the lowest perplexity score viewed as the most probable and reliable trial path given the model's capacity. The trial number is negatively correlated with the prediction confidence, i.e., the less confident the model is, the more trials it should sample. This is consistent with human beings' behaviour: when feeling uncertain or unconfident, one tends to think more creatively, exploring multiple thinking paths, to cautiously select the path one feels most confident about. Extensive experiments on both LLMs and MLLMs show that our proposed CNTP approach outperforms existing standard decoding strategies consistently by a clear margin. Moreover, the integration of CNTP with self consistency can further improve over vanilla self consistency. We believe our proposed CNTP has the potential to become one of the default choices for LLM decoding. Code is available at https://github.com/wyzjack/CNTP.

View Paper

Fairness Evaluation of Large Language Models in Academic Library Reference Services

Authors: Haining Wang, Jason Clark, Yueru Yan, Star Bradley, Ruiyang Chen, Yiqiong Zhang, Hengyi Fu, Zuoyu Tian

Abstract: arXiv:2507.04224v2 Announce Type: replace-cross Abstract: As libraries explore large language models (LLMs) for use in virtual reference services, a key question arises: Can LLMs serve all users equitably, regardless of demographics or social status? While they offer great potential for scalable support, LLMs may also reproduce societal biases embedded in their training data, risking the integrity of libraries' commitment to equitable service. To address this concern, we evaluate whether LLMs differentiate responses across user identities by prompting six state-of-the-art LLMs to assist patrons differing in sex, race/ethnicity, and institutional role. We found no evidence of differentiation by race or ethnicity, and only minor evidence of stereotypical bias against women in one model. LLMs demonstrated nuanced accommodation of institutional roles through the use of linguistic choices related to formality, politeness, and domain-specific vocabularies, reflecting professional norms rather than discriminatory treatment. These findings suggest that current LLMs show a promising degree of readiness to support equitable and contextually appropriate communication in academic library reference services.

View Paper

Infinite Video Understanding

Authors: Dell Zhang, Xiangyu Chen, Jixiang Luo, Mengxi Jia, Changzhi Sun, Ruilong Ren, Jingren Liu, Hao Sun, Xuelong Li

Abstract: arXiv:2507.09068v2 Announce Type: replace-cross Abstract: The rapid advancements in Large Language Models (LLMs) and their multimodal extensions (MLLMs) have ushered in remarkable progress in video understanding. However, a fundamental challenge persists: effectively processing and comprehending video content that extends beyond minutes or hours. While recent efforts like Video-XL-2 have demonstrated novel architectural solutions for extreme efficiency, and advancements in positional encoding such as HoPE and VideoRoPE++ aim to improve spatio-temporal understanding over extensive contexts, current state-of-the-art models still encounter significant computational and memory constraints when faced with the sheer volume of visual tokens from lengthy sequences. Furthermore, maintaining temporal coherence, tracking complex events, and preserving fine-grained details over extended periods remain formidable hurdles, despite progress in agentic reasoning systems like Deep Video Discovery. This position paper posits that a logical, albeit ambitious, next frontier for multimedia research is Infinite Video Understanding -- the capability for models to continuously process, understand, and reason about video data of arbitrary, potentially never-ending duration. We argue that framing Infinite Video Understanding as a blue-sky research objective provides a vital north star for the multimedia, and the wider AI, research communities, driving innovation in areas such as streaming architectures, persistent memory mechanisms, hierarchical and adaptive representations, event-centric reasoning, and novel evaluation paradigms. Drawing inspiration from recent work on long/

ultra-long video understanding and several closely related fields, we outline the core challenges and key research directions towards achieving this transformative capability.

View Paper

Advanced U-Net Architectures with CNN Backbones for Automated Lung Cancer Detection and Segmentation in Chest CT Images

Authors: Alireza Golkarieh, Kiana Kiashemshaki, Sajjad Rezvani Boroujeni, Nasibeh Asadi Isakan

Abstract: arXiv:2507.09898v2 Announce Type: replace-cross Abstract: This study investigates the effectiveness of U-Net architectures integrated with various convolutional neural network (CNN) backbones for automated lung cancer detection and segmentation in chest CT images, addressing the critical need for accurate diagnostic tools in clinical settings. A balanced dataset of 832 chest CT images (416 cancerous and 416 non-cancerous) was preprocessed using Contrast Limited Adaptive Histogram Equalization (CLAHE) and resized to 128x128 pixels. U-Net models were developed with three CNN backbones: ResNet50, VGG16, and Xception, to segment lung regions. After segmentation, CNN-based classifiers and hybrid models combining CNN feature extraction with traditional machine learning classifiers (Support Vector Machine, Random Forest, and Gradient Boosting) were evaluated using 5-fold cross-validation. Metrics included accuracy, precision, recall, F1-score, Dice coefficient, and ROC-AUC. U-Net with ResNet50 achieved the best performance for cancerous lungs (Dice: 0.9495, Accuracy: 0.9735), while U-Net with VGG16 performed best for non-cancerous segmentation (Dice: 0.9532, Accuracy: 0.9513). For classification, the CNN model using U-Net with Xception achieved 99.1 percent accuracy, 99.74 percent recall, and 99.42 percent F1-score. The hybrid CNN-SVM-Xception model achieved 96.7 percent accuracy and 97.88 percent F1-score. Compared to prior methods, our framework consistently outperformed existing models. In conclusion, combining U-Net with advanced CNN backbones provides a powerful method for both segmentation and classification of lung cancer in CT scans, supporting early diagnosis and clinical decision-making.

View Paper

SToFM: a Multi-scale Foundation Model for Spatial Transcriptomics

Authors: Suyuan Zhao, Yizhen Luo, Ganbo Yang, Yan Zhong, Hao Zhou, Zaiqing Nie

Abstract: arXiv:2507.11588v2 Announce Type: replace-cross Abstract: Spatial Transcriptomics (ST) technologies provide biologists with rich insights into single-cell biology by preserving spatial context of cells. Building foundational models for ST can significantly enhance the analysis of vast and complex data sources, unlocking new perspectives on the intricacies of biological tissues. However, modeling ST data is inherently challenging due to the need to extract multi-scale information from tissue slices containing vast numbers of cells. This process requires integrating macro-scale tissue morphology, micro-scale cellular microenvironment, and gene-scale gene expression profile. To address this challenge, we propose SToFM, a multi-scale Spatial Transcriptomics Foundation Model. SToFM first performs multi-scale information extraction on each ST slice, to construct a set of ST sub-slices that aggregate macro-, micro- and gene-scale information. Then an SE(2) Transformer is used to obtain high-quality cell representations from the sub-slices. Additionally, we construct \textbf{SToCorpus-88M}, the largest high-resolution spatial transcriptomics corpus for pretraining. SToFM achieves outstanding performance on a variety of downstream tasks, such as tissue region semantic segmentation and cell type annotation, demonstrating its comprehensive understanding of ST data through capturing and integrating multi-scale information.

View Paper

Fragment size density estimator for shrinkage-induced fracture based on a physics-informed neural network

Authors: Shin-ichi Ito

Abstract: arXiv:2507.11799v2 Announce Type: replace-cross Abstract: This paper presents a neural network (NN)-based solver for an integro-differential equation that models shrinkage-induced fragmentation. The proposed method directly maps input parameters to the corresponding probability density function without numerically solving the governing equation, thereby significantly reducing computational costs. Specifically, it enables efficient evaluation of the density function in Monte Carlo simulations while maintaining accuracy comparable to or even exceeding that of conventional finite difference schemes. Validatation on synthetic data demonstrates both the method's computational efficiency and predictive reliability. This study establishes a foundation for the data-driven inverse analysis of fragmentation and suggests the potential for extending the framework beyond pre-specified model structures.

View Paper

Spatial Frequency Modulation for Semantic Segmentation

Authors: Linwei Chen, Ying Fu, Lin Gu, Dezhi Zheng, Jifeng Dai

Abstract: arXiv:2507.11893v2 Announce Type: replace-cross Abstract: High spatial frequency information, including fine details like textures, significantly contributes to the accuracy of semantic segmentation. However, according to the Nyquist-Shannon Sampling Theorem, high-frequency components are vulnerable to aliasing or distortion when propagating through downsampling layers such as strided-convolution. Here, we propose a novel Spatial Frequency Modulation (SFM) that modulates high-frequency features to a lower frequency before downsampling and then demodulates them back during upsampling. Specifically, we implement modulation through adaptive resampling (ARS) and design a lightweight add-on that can densely sample the high-frequency areas to scale up the signal, thereby lowering its frequency in accordance with the Frequency Scaling Property. We also propose Multi-Scale Adaptive Upsampling (MSAU) to demodulate the modulated feature and recover high-frequency information through non-uniform upsampling This module further improves segmentation by explicitly exploiting information interaction between densely and sparsely resampled areas at multiple scales. Both modules can seamlessly integrate with various architectures, extending from convolutional neural networks to transformers. Feature visualization and analysis confirm that our method effectively alleviates aliasing while successfully retaining details after demodulation. Finally, we validate the broad applicability and effectiveness of SFM by extending it to image classification, adversarial robustness, instance segmentation, and panoptic segmentation tasks. The code is available at https://github.com/Linwei-Chen/SFM.

View Paper

Frequency-Dynamic Attention Modulation for Dense Prediction

Authors: Linwei Chen, Lin Gu, Ying Fu

Abstract: arXiv:2507.12006v2 Announce Type: replace-cross Abstract: Vision Transformers (ViTs) have significantly advanced computer vision, demonstrating strong performance across various tasks. However, the attention mechanism in ViTs makes each layer function as a low-pass filter, and the stacked-layer architecture in existing transformers suffers from frequency vanishing. This leads to the loss of critical details and textures. We propose a novel, circuit-theory-inspired strategy called Frequency-Dynamic Attention Modulation (FDAM), which can be easily plugged into ViTs. FDAM directly modulates the overall frequency response of ViTs and consists of two techniques: Attention Inversion (AttInv) and Frequency Dynamic Scaling (FreqScale). Since circuit theory uses low-pass filters as fundamental elements, we introduce AttInv, a method that generates complementary high-pass filtering by inverting the low-pass filter in the attention matrix, and dynamically combining the two. We further design FreqScale to weight different frequency components for fine-grained adjustments to the target response function. Through feature similarity analysis and effective rank evaluation, we demonstrate that our approach avoids representation collapse, leading to consistent performance improvements across various models, including SegFormer, DeiT, and MaskDINO. These improvements are evident in tasks such as semantic segmentation, object detection, and instance segmentation. Additionally, we apply our method to remote sensing detection, achieving state-of-the-art results in single-scale settings. The code is available at https://github.com/Linwei-Chen/FDAM.

View Paper

Feature-Enhanced TResNet for Fine-Grained Food Image Classification

Authors: Lulu Liu, Zhiyong Xiao

Abstract: arXiv:2507.12828v2 Announce Type: replace-cross Abstract: Food is not only essential to human health but also serves as a medium for cultural identity and emotional connection. In the context of precision nutrition, accurately identifying and classifying food images is critical for dietary monitoring, nutrient estimation, and personalized health management. However, fine-grained food classification remains challenging due to the subtle visual differences among similar dishes. To address this, we propose Feature-Enhanced TResNet (FE-TResNet), a novel deep learning model designed to improve the accuracy of food image recognition in fine-grained scenarios. Built on the TResNet architecture, FE-TResNet integrates a Style-based Recalibration Module (StyleRM) and Deep Channel-wise Attention (DCA) to enhance feature extraction and emphasize subtle distinctions between food items. Evaluated on two benchmark Chinese food datasets-ChineseFoodNet and CNFOOD-241-FE-TResNet achieved high classification accuracies of 81.37% and 80.29%, respectively. These results demonstrate its effectiveness and highlight its potential as a key enabler for intelligent dietary assessment and personalized recommendations in precision nutrition systems.

View Paper

Photonic Fabric Platform for AI Accelerators

Authors: Jing Ding, Trung Diep

Abstract: arXiv:2507.14000v3 Announce Type: replace-cross Abstract: This paper presents the Photonic FabricTM and the Photonic Fabric ApplianceTM (PFA), a photonic-enabled switch and memory subsystem that delivers low latency, high bandwidth, and low per-bit energy. By integrating high-bandwidth HBM3E memory, an on-module photonic switch, and external DDR5 in a 2.5D electro-optical system-in-package, the PFA offers up to 32 TB of shared memory alongside 115 Tbps of all-to-all digital switching. The Photonic FabricTM enables distributed AI training and inference to execute parallelism strategies more efficiently. The Photonic Fabric removes the silicon beachfront constraint that limits the fixed memory-to-compute ratio observed in virtually all current XPU accelerator designs. Replacing a local HBM stack on an XPU with a chiplet that connects to the Photonic Fabric increases its memory capacity and correspondingly its memory bandwidth by offering a flexible path to scaling well beyond the limitations of on-package HBM alone. We introduce CelestiSim, a lightweight analytical simulator validated on NVIDIA H100 and H200 systems. It is used to evaluate the performance of LLM reference and energy savings on PFA, without any significant change to the GPU core design. With the PFA, the simulation results show that up to 3.66x throughput and 1.40x latency improvements in LLM inference at 405B parameters, up to 7.04x throughput and 1.41x latency improvements at 1T parameters, and 60-90% energy savings in data movement for heavy collective operations in all LLM training scenarios. While these results are shown for NVIDIA GPUs, they can be applied similarly to other AI accelerator designs (XPUs) that share the same fundamental limitation of fixed memory to compute.

View Paper

NeuroHD-RA: Neural-distilled Hyperdimensional Model with Rhythm Alignment

Authors: ZhengXiao He, Jinghao Wen, Huayu Li, Siyuan Tian, Ao Li

Abstract: arXiv:2507.14184v3 Announce Type: replace-cross Abstract: We present a novel and interpretable framework for electrocardiogram (ECG)-based disease detection that combines hyperdimensional computing (HDC) with learnable

neural encoding. Unlike conventional HDC approaches that rely on static, random projections, our method introduces a rhythm-aware and trainable encoding pipeline based on RR intervals, a physiological signal segmentation strategy that aligns with cardiac cycles. The core of our design is a neural-distilled HDC architecture, featuring a learnable RR-block encoder and a BinaryLinear hyperdimensional projection layer, optimized jointly with cross-entropy and proxy-based metric loss. This hybrid framework preserves the symbolic interpretability of HDC while enabling task-adaptive representation learning. Experiments on Apnea-ECG and PTB-XL demonstrate that our model significantly outperforms traditional HDC and classical ML baselines, achieving 73.09\% precision and an F1 score of 0.626 on Apnea-ECG, with comparable robustness on PTB-XL. Our framework offers an efficient and scalable solution for edge-compatible ECG classification, with strong potential for interpretable and personalized health monitoring.

View Paper

Artificial Intelligence for Green Hydrogen Yield Prediction and Site Suitability using SHAP-Based Composite Index: Focus on Oman

Authors: Obumneme Zimuzor Nwafor, Mohammed Abdul Majeed Al Hooti

Abstract: arXiv:2507.14219v2 Announce Type: replace-cross Abstract: As nations seek sustainable alternatives to fossil fuels, green hydrogen has emerged as a promising strategic pathway toward decarbonisation, particularly in solar-rich arid regions. However, identifying optimal locations for hydrogen production requires the integration of complex environmental, atmospheric, and infrastructural factors, often compounded by limited availability of direct hydrogen yield data. This study presents a novel Artificial Intelligence (AI) framework for computing green hydrogen yield and site suitability index using mean absolute SHAP (SHapley Additive exPlanations) values. This framework consists of a multi-stage pipeline of unsupervised multi-variable clustering, supervised machine learning classifier and SHAP algorithm. The pipeline trains on an integrated meteorological, topographic and temporal dataset and the results revealed distinct spatial patterns of suitability and relative influence of the variables. With model predictive accuracy of 98%, the result also showed that water proximity, elevation and seasonal variation are the most influential factors determining green hydrogen site suitability in Oman with mean absolute shap values of 2.470891, 2.376296 and 1.273216 respectively. Given limited or absence of ground-truth yield data in many countries that have green hydrogen prospects and ambitions, this study offers an objective and reproducible alternative to subjective expert weightings, thus allowing the data to speak for itself and

potentially discover novel latent groupings without pre-imposed assumptions. This study offers industry stakeholders and policymakers a replicable and scalable tool for green hydrogen infrastructure planning and other decision making in data-scarce regions.

View Paper

APTx Neuron: A Unified Trainable Neuron Architecture Integrating Activation and Computation

Authors: Ravin Kumar

Abstract: arXiv:2507.14270v2 Announce Type: replace-cross Abstract: We propose the APTx Neuron, a novel, unified neural computation unit that integrates non-linear activation and linear transformation into a single trainable expression. The APTx Neuron is derived from the APTx activation function, thereby eliminating the need for separate activation layers and making the architecture both computationally efficient and elegant. The proposed neuron follows the functional form $y = \sum*{i=1}^{n} ((\|alpha*i + \tanh(\beta*i x*i)) \cdot \gamma*i x*i) + \delta$, where all parameters $\alpha*i$, $\|beta*i$, γ_i, and $ \delta$ are trainable. We validate our APTx Neuron-based architecture on the MNIST dataset, achieving up to 96.69% test accuracy in just 20 epochs using approximately 332K trainable parameters. The results highlight the superior expressiveness and computational efficiency of the APTx Neuron compared to traditional neurons, pointing toward a new paradigm in unified neuron design and the architectures built upon it.

View Paper

SegQuant: A Semantics-Aware and Generalizable Quantization Framework for Diffusion Models

Authors: Jiaji Zhang, Ruichao Sun, Hailiang Zhao, Jiaju Wu, Peng Chen, Hao Li, Yuying Liu, Xinkui Zhao, Kingsum Chow, Gang Xiong, Shuiguang Deng

Abstract: arXiv:2507.14811v2 Announce Type: replace-cross Abstract: Diffusion models have demonstrated exceptional generative capabilities but are computationally intensive, posing significant challenges for deployment in resource-constrained or latency-sensitive environments. Quantization offers an effective means to reduce model size and computational cost, with post-training quantization (PTQ) being particularly appealing due to its compatibility with pre-trained models without requiring retraining or training data. However, existing

PTQ methods for diffusion models often rely on architecture-specific heuristics that limit their generalizability and hinder integration with industrial deployment pipelines. To address these limitations, we propose SegQuant, a unified quantization framework that adaptively combines complementary techniques to enhance cross-model versatility. SegQuant consists of a segment-aware, graph-based quantization strategy (SegLinear) that captures structural semantics and spatial heterogeneity, along with a dual-scale quantization scheme (DualScale) that preserves polarity-asymmetric activations, which is crucial for maintaining visual fidelity in generated outputs. SegQuant is broadly applicable beyond Transformer-based diffusion models, achieving strong performance while ensuring seamless compatibility with mainstream deployment tools.

View Paper

EndoControlMag: Robust Endoscopic Vascular Motion Magnification with Periodic Reference Resetting and Hierarchical Tissue-aware Dual-Mask Contro

Authors: An Wang, Rulin Zhou, Mengya Xu, Yiru Ye, Longfei Gou, Yiting Chang, Hao Chen, Chwee Ming Lim, Jiankun Wang, Hongliang Ren

Abstract: arXiv:2507.15292v3 Announce Type: replace-cross Abstract: Visualizing subtle vascular motions in endoscopic surgery is crucial for surgical precision and decision-making, yet remains challenging due to the complex and dynamic nature of surgical scenes. To address this, we introduce EndoControlMag, a training-free, Lagrangian-based framework with mask-conditioned vascular motion magnification tailored to endoscopic environments. Our approach features two key modules: a Periodic Reference Resetting (PRR) scheme that divides videos into short overlapping clips with dynamically updated reference frames to prevent error accumulation while maintaining temporal coherence, and a Hierarchical Tissue-aware Magnification (HTM) framework with dual-mode mask dilation. HTM first tracks vessel cores using a pretrained visual tracking model to maintain accurate localization despite occlusions and view changes. It then applies one of two adaptive softening strategies to surrounding tissues: motion-based softening that modulates magnification strength proportional to observed tissue displacement, or distance-based exponential decay that simulates biomechanical force attenuation. This dual-mode approach accommodates diverse surgical scenarios-motion-based softening excels with complex tissue deformations while distance-based softening provides stability during unreliable optical flow conditions. We evaluate EndoControlMag on our EndoVMM24 dataset spanning four different surgery types and various challenging scenarios, including occlusions, instrument disturbance, view

changes, and vessel deformations. Quantitative metrics, visual assessments, and expert surgeon evaluations demonstrate that EndoControlMag significantly outperforms existing methods in both magnification accuracy and visual quality while maintaining robustness across challenging surgical conditions. The code, dataset, and video results are available at https://szupc.github.io/EndoControlMag/.

View Paper

Solving nonconvex Hamilton--Jacobi--Isaacs equations with PINN-based policy iteration

Authors: Hee Jun Yang, Minjung Gim, Yeoneung Kim

Abstract: arXiv:2507.15455v2 Announce Type: replace-cross Abstract: We propose a mesh-free policy iteration framework that combines classical dynamic programming with physics-informed neural networks (PINNs) to solve high-dimensional, nonconvex Hamilton--Jacobi--Isaacs (HJI) equations arising in stochastic differential games and robust control. The method alternates between solving linear second-order PDEs under fixed feedback policies and updating the controls via pointwise minimax optimization using automatic differentiation. Under standard Lipschitz and uniform ellipticity assumptions, we prove that the value function iterates converge locally uniformly to the unique viscosity solution of the HJI equation. The analysis establishes equi-Lipschitz regularity of the iterates, enabling provable stability and convergence without requiring convexity of the Hamiltonian. Numerical experiments demonstrate the accuracy and scalability of the method. In a two-dimensional stochastic path-planning game with a moving obstacle, our method matches finite-difference benchmarks with relative L^2-errors below %10^{-2}%. In five- and ten-dimensional publisher-subscriber differential games with anisotropic noise, the proposed approach consistently outperforms direct PINN solvers, yielding smoother value functions and lower residuals. Our results suggest that integrating PINNs with policy iteration is a practical and theoretically grounded method for solving high-dimensional, nonconvex HJI equations, with potential applications in robotics, finance, and multi-agent reinforcement learning.

View Paper

SFNet: A Spatial-Frequency Domain Deep Learning Network for Efficient Alzheimer's Disease Diagnosis

Authors: Xinyue Yang, Meiliang Liu, Yunfang Xu, Xiaoxiao Yang, Zhengye Si, Zijin Li, Zhiwen Zhao

Abstract: arXiv:2507.16267v2 Announce Type: replace-cross Abstract: Alzheimer's disease (AD) is a progressive neurodegenerative disorder that predominantly affects the elderly population and currently has no cure. Magnetic Resonance Imaging (MRI), as a non-invasive imaging technique, is essential for the early diagnosis of AD. MRI inherently contains both spatial and frequency information, as raw signals are acquired in the frequency domain and reconstructed into spatial images via the Fourier transform. However, most existing AD diagnostic models extract features from a single domain, limiting their capacity to fully capture the complex neuroimaging characteristics of the disease. While some studies have combined spatial and frequency information, they are mostly confined to 2D MRI, leaving the potential of dual-domain analysis in 3D MRI unexplored. To overcome this limitation, we propose Spatio-Frequency Network (SFNet), the first end-to-end deep learning framework that simultaneously leverages spatial and frequency domain information to enhance 3D MRI-based AD diagnosis. SFNet integrates an enhanced dense convolutional network to extract local spatial features and a global frequency module to capture global frequency-domain representations. Additionally, a novel multi-scale attention module is proposed to further refine spatial feature extraction. Experiments on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset demonstrate that SFNet outperforms existing baselines and reduces computational overhead in classifying cognitively normal (CN) and AD, achieving an accuracy of 95.1%.

View Paper

EarthCrafter: Scalable 3D Earth Generation via Dual-Sparse Latent Diffusion

Authors: Shang Liu, Chenjie Cao, Chaohui Yu, Wen Qian, Jing Wang, Fan Wang

Abstract: arXiv:2507.16535v2 Announce Type: replace-cross Abstract: Despite the remarkable developments achieved by recent 3D generation works, scaling these methods to geographic extents, such as modeling thousands of square kilometers of Earth's surface, remains an open challenge. We address this through a dual innovation in data infrastructure and model architecture. First, we introduce Aerial-Earth3D, the largest 3D aerial dataset to date, consisting of 50k curated scenes (each measuring 600m x 600m) captured across the U.S. mainland, comprising 45M multi-view Google Earth frames. Each scene provides pose-annotated multi-view images, depth maps, normals, semantic segmentation, and camera poses, with explicit quality control to ensure terrain diversity. Building on this foundation, we propose EarthCrafter, a tailored framework for large-scale 3D Earth generation via sparse-decoupled latent diffusion. Our architecture separates structural and textural generation: 1) Dual sparse 3D-VAEs compress high-resolution geometric voxels and textural 2D Gaussian Splats (2DGS) into compact latent spaces, largely alleviating the costly computation suffering from vast geographic scales while preserving critical information. 2) We propose condition-aware flow matching models trained on mixed inputs (semantics, images, or neither) to flexibly model latent geometry and texture features independently. Extensive experiments demonstrate that EarthCrafter performs substantially better in extremely large-scale generation. The framework further supports versatile applications, from semantic-guided urban layout generation to unconditional terrain synthesis, while maintaining geographic plausibility through our rich data priors from Aerial-Earth3D. Our project page is available at https://whiteinblue.github.io/earthcrafter/

View Paper