

The 1st Workshop for Research on Agent Language Models (REALMS 2025)

Vienna, Austria
31 July 2025

ISBN: 979-8-3313-2415-5

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>Prompt-based Personality Profiling: Reinforcement Learning for Relevance Filtering</i> Jan Hofmann, Cornelia Sindermann and Roman Klinger	1
<i>DFLOW: Diverse Dialogue Flow Simulation with Large Language Models</i> Wanyu Du, Song Feng, James Gung, Lijia Sun, Yi Zhang, Saab Mansour and Yanjun Qi	17
<i>CAMPHOR: Collaborative Agents for Multi-input Planning and High-Order Reasoning On Device</i> Yicheng Fu, Raviteja Anantha and Jianpeng Cheng	33
<i>A Multi-AI Agent System for Autonomous Optimization of Agentic AI Solutions via Iterative Refinement and LLM-Driven Feedback Loops</i> Kamer Ali Yuksel, Thiago Castro Ferreira, Mohamed Al-Badrashiny and Hassan Sawaf	52
<i>The Art of Tool Interface Design</i> Yunnan Wu, Qile P. Chen, Deshank Baranwal, Jinlong Zhou and Jian Yuan	63
<i>AID-Agent: An LLM-Agent for Advanced Extraction and Integration of Documents</i> Bin Li, Jannis Conen and Felix Aller	80
<i>Hidden Forms: A Dataset to Fill Masked Interfaces from Language Commands</i> Anirudh Sundar, Christopher Gordon Richardson, William Gay, Benjamin Reichman and Larry Heck	89
<i>Do Large Language Models Learn Human-Like Strategic Preferences?</i> Jesse Roberts, Kyle Moore and Douglas Fisher	97
<i>Inherent and emergent liability issues in LLM-based agentic systems: a principal-agent perspective</i> Garry A. Gabison and R. Patrick Xian	109
<i>Positive Experience Reflection for Agents in Interactive Text Environments</i> Philip Lippmann, Matthijs T. J. Spaan and Jie Yang	131
<i>PAARS: Persona Aligned Agentic Retail Shoppers</i> Saab Mansour, Leonardo Perelli, Lorenzo Mainetti, George Davidson and Stefano D’Amato	143
<i>Leveraging LLM-based sentiment analysis for portfolio optimization with proximal policy optimization</i> Kemal Kirtac and Guido Germano	160
<i>Safe in Isolation, Dangerous Together: Agent-Driven Multi-Turn Decomposition Jailbreaks on LLMs</i> Devansh Srivastav and Xiao Zhang	170
<i>ToolReflection: Improving Large Language Models for Real-World API Calls with Self-Generated Data</i> Gregory Polyakov, Ilseyar Alimova, Dmitry Abulkhanov, Ivan Sedykh, Andrey Bout, Sergey Nikolenko and Irina Piontkovskaya	184
<i>Conditional Multi-Stage Failure Recovery for Embodied Agents</i> Younna Farag, Svetlana Stoyanchev, Mohan Li, Simon Keizer and Rama Doddipatla	200
<i>Snap Out of It: A Dual-Process Approach to Mitigating Overthinking in Language Model Reasoning</i> Ashish Pandian, Nelson Lojo, Wei Xun Lai and Jackson Lukas	228
<i>A Conversational Agent Framework for Multimodal Knowledge Retrieval: A Case Study in FHWA Info-Highway Web Portal Queries</i> Sai Surya Gadiraju, Duoduo Liao and Zijie He	250

<i>A Study on Leveraging Search and Self-Feedback for Agent Reasoning</i>	
Karthikeyan K, Michelle Yuan, Elman Mansimov, Katerina Margatina, Anurag Pratik, Daniele Bonadiman, Monica Sunkara, Yi Zhang and Yassine Benajiba	259
<i>GitGoodBench: A Novel Benchmark For Evaluating Agentic Performance On Git</i>	
Tobias Lindenbauer, Egor Bogomolov and Yaroslav Zharov	272
<i>TCQA²: A Tiered Conversational Q&A Agent in Gaming</i>	
Ze Chen, Chengcheng Wei, Jiewen Zheng and Jiarong He	289
<i>Oversight Structures for Agentic AI in Public-Sector Organizations</i>	
Chris Schmitz, Jonathan Rystrom and Jan Batzner	298
<i>Are You Sure You're Positive? Consolidating Chain-of-Thought Agents with Uncertainty Quantification for Aspect-Category Sentiment Analysis</i>	
Filippos Ventirozos, Peter A. Appleby and Matthew Shardlow	309
<i>Bridging the Digital Divide: Empowering Elderly Smartphone Users with Intelligent and Human-Centered Design in Agemate</i>	
Liangliang Chen and Yongzhen Mu	327
<i>Decentralized Low-Rank Fine-Tuning of Large Language Models</i>	
Sajjad Ghiasvand, Mahnoosh Alizadeh and Ramtin Pedarsani	334
<i>Measuring temporal effects of agent knowledge by date-controlled tool use</i>	
R. Patrick Xian, Qiming Cui, Stefan Bauer and Reza Abbasi-Asl	346
<i>VisTRA: Visual Tool-use Reasoning Analyzer for Small Object Visual Question Answering</i>	
Hiroaki Sugiyama, Ko Koga and Toshifumi Nishijima	356
<i>StateAct: Enhancing LLM Base Agents via Self-prompting and State-tracking</i>	
Nikolai Rozanov and Marek Rei	367
<i>DIAMOND: An LLM-Driven Agent for Context-Aware Baseball Highlight Summarization</i>	
Jeonghun Kang, Soonmok Kwon, Joonseok Lee and Byung-Hak Kim	386
<i>RL + Transformer = A General-Purpose Problem Solver</i>	
Micah Rentschler and Jesse Roberts	401
<i>From Knowledge to Noise: CTIM-Rover and the Pitfalls of Episodic Memory in Software Engineering Agents</i>	
Tobias Lindenbauer, Georg Groh and Hinrich Schuetze	411
<i>FrontierScience Bench: Evaluating AI Research Capabilities in LLMs</i>	
Matthew Li, Santiago Torres-Garcia, Shayan Halder, Phani Kuppa, Sean O'Brien, Vasu Sharma, Kevin Zhu and Sunishchal Dev	428
<i>The Power of Simplicity in LLM-Based Event Forecasting</i>	
Meiru Zhang, Auss Abbood, Zaiqiao Meng and Nigel Collier	454
<i>Weight-of-Thought Reasoning: Exploring Neural Network Weights for Enhanced LLM Reasoning</i>	
Saif Punjwani and Larry Heck	471