

Wordplay 2025

Wordplay: When Language meets Games

Proceedings of the Workshop

November 9, 2025

The Wordplay organizers gratefully acknowledge the support from the following sponsors.

Gold



In cooperation with



©2025 Empirical Methods in Natural Language Processing (EMNLP)

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN N/A

Introduction

This workshop will focus on exploring the utility of interactive narratives, think everything from classic text-adventures like Zork to modern Twine games, to fill a role as the learning environments of choice for language-based tasks including but not limited to storytelling. A few previous iterations of this workshop took place very successfully with hundreds of attendees, at NeurIPS 2018, NeurIPS 2020, NAACL 2022, and ACL 2024. Since then, the community of people working in this area has rapidly increased. This workshop aims to be a centralized place where all researchers involved across a breadth of fields can interact and learn from each other. Furthermore, it will act as a showcase to the wider NLP/RL/Game communities on interactive narrative's place as a learning environment. The program will feature a collection of invited talks in addition to contributed talks and posters from each of these sections of the interactive narrative community and the wider NLP and RL communities.

We like all things:

- Interactive narrative: game playing RL agents, game generation, etc.
- Interactive language learning
- Natural language generation
- Improvisational storytelling
- And more! Anything you can think of that involves narrative, interactivity, and language!

Full details about the workshop, including the keynotes, accepted papers, and the schedule can be found on the workshop website:

<https://wordplay-workshop.github.io/>

Organizing Committee

Organizers

Xingdi Yuan, Microsoft Research

Hiromi Wakaki, Sony Group Corporation

Lara Martin, University of Maryland, Baltimore County (UBMC)

Antoine Bosselut, École Polytechnique Fédérale de Lausanne (EPFL)

Ruiyi Wang, University of California, San Diego (UCSD)

Ruyao Wang, Central University of Finance and Economics

Marc-Alexandre Côté, Microsoft Research

Prithviraj Ammanabrolu, University of California, San Diego (UCSD) & Nvidia

Program Committee

Program Committee

Anchao Li, ByteDance

Arya Honraopatil, University of Maryland, Baltimore County

Ashutosh Adhikari, University of Edinburgh

Christopher Zhang Cui, University of California, San Diego

Hosein Rezaei, University of York

Kuan Lu, Cornell University

Marcus McAllister, University of Maryland, Baltimore County

Mete Ismayilzada, Swiss Federal Technology Institute of Lausanne

Nischal Reddy Chandra, Adobe

Patricia J. Delafuente, University of Maryland, Baltimore County

Samuel Barry, Mistral AI

Shadab Hafiz Choudhury, University of Maryland, Baltimore County

Silin Gao, Swiss Federal Technology Institute of Lausanne

Keynote Talk

Beyond Exams: Evaluating LLM Agents Through Interactive Collaboration and Competition

Diyi Yang
Stanford University

Abstract: TBD

Bio: Diyi Yang is an assistant professor in the Computer Science Department at Stanford University, also affiliated with the Stanford NLP Group, Stanford HCI Group and Stanford Human Centered AI Institute. Her research focuses on human-centered natural language processing and human-AI interaction. She is a recipient of Microsoft Research Faculty Fellowship (2021), NSF CAREER Award (2022), and a Sloan Research Fellowship (2024). Her work has received multiple paper awards or nominations at top NLP and HCI conferences.

Keynote Talk

Keynote 2

Sudha Rao
Microsoft

Abstract: In our team at Microsoft Research, we have been exploring how generative AI can transform narrative design games. In the first part of the talk, I will introduce GENEVA, a prototype tool for generating and visualizing branching narratives using large language models (LLMs). This tool allows narrative designers to develop and iterate narratives much faster than traditional tools. In the second part, I will present a case study on how generative AI can evolve game narratives in real-time based on player feedback. The goal of this work is to understand how players, game creators and generative AI can collaboratively create and evolve a game’s narrative over time. Learn more about our team’s work on our website – <https://msr-emergence.com/>.

Bio: Sudha Rao is a Principal Researcher in the Natural Language Processing group at Microsoft Research. Currently, her work focuses on applying large language models to narrative-rich games. Particularly, she is interested in understanding how players, game creators and generative AI can collaboratively evolve a game’s narrative over time. Our group’s recent work on game narrative graphs was crucial in fostering XBOX’s partnership with Inworld AI. She received her Ph.D. in Computer Science from University of Maryland, College Park under the guidance of Dr. Hal Daume III. Her thesis work was on generating clarification questions won the best paper award at ACL 2018.

Keynote Talk

Keynote 3

Xin Wang

University of California Santa Barbara (USCB) & Simular

Abstract: TBD

Bio: Xin (Eric) Wang is an Assistant Professor in the Computer Science Department at UC Santa Barbara and the Head of Research at Simular. He also directs the UCSB Center for Responsible Machine Learning (CRML). His research interests include Natural Language Processing, Computer Vision, and Machine Learning, with an emphasis on Multimodal and Agentic AI. Previously, he was a faculty at UC Santa Cruz and also spent time at Google Research, Meta FAIR, Microsoft Research, and Adobe Research. Eric has served as (Senior) Area Chair for conferences such as ACL, NAACL, EMNLP, ICLR, and NeurIPS, and organized workshops and tutorials at those venues. He has received several awards and recognitions for his work, including Best Paper Awards from CVPR and ICLRW, Google Research Faculty Award, Amazon Alexa Prize Awards, JPMorganChase Faculty Research Award, Cisco Faculty Research Award, eBay Faculty Research Awards, AAIL Interdisciplinary Research Award, and various gift awards/grants from Adobe, Apple, Snap, Microsoft, OpenAI, Cybever, Orby, etc.

Keynote Talk

Keynote 4

Boyang Li

Nanyang Technological University (NTU)

Abstract: TBD

Bio: Boyang Li is an Associate Professor at the College of Computing and Data Science, Nanyang Technological University, Singapore. His research interests lie broadly in Multi-modal Learning, Data-centric AI, Machine Learning, and Computational Narrative Intelligence. Some notable works of his in multimodal learning include InstructBLIP, Plug-and-Play VQA, and VisualGPT. His work in Data-centric AI covers the understanding of datasets (factor analysis of datasets), data influence to model decisions (HyDRA), selection of training data (COINCIDE), and synthesis of training data from generative AI (SPARCL). Another work direction is to build AI that understands commonsense causal relations between events. Examples include heavy sweating causes thirst and hearing compliments makes people happy". Such causal conditions are neither necessary nor sufficient, as they are context-dependent. However, they are indispensable to a world model that enables AI to interpret the past and to plan for the future. Recent representative work includes benchmarks (BlackSwanSuite and M-SyMoN) and a technique to extract such knowledge from LLMs, though his work in this direction traces back to his PhD work on learning event graphs. Prior to NTU, he was a Senior Research Scientist at Baidu Research USA, and a Research Scientist and Group Leader at Disney Research. He received his Ph.D. degree from Georgia Institute of Technology.

Keynote Talk

LLM Agents for In-the-Wild Video Game-Playing

Jaewoo Ahn

Seoul National University (SNU)

Abstract: Large Language Model (LLM) powered agents, including Vision Language Model (VLM) agents, are revolutionizing digital interaction, particularly within the complex, dynamic environments of in-the-wild video games. However, a significant gap persists between current academic benchmarks and the demands of real-world commercial titles. Existing benchmarks often lack genre diversity, fail to evaluate agents on long-horizon, narrative-driven tasks, and overlook the practical challenges of deploying agents in live games. This talk addresses these critical gaps by introducing a two-part approach to building and evaluating robust gaming agents. First, I will present Orak, a foundational benchmark featuring 12 diverse, popular video games designed to comprehensively assess LLM/VLM capabilities and agentic modules required for generalist gameplay. Second, I will introduce FlashAdventure, a benchmark of 34 Flash-based adventure games (e.g., room escape, mystery/detective, visual novel, etc) focused on the challenge of full story arc completion, tackling the observation-behavior gap with a novel agentic framework, COAST, which leverages long-term cluememory to better plan and solve sequential tasks. Finally, I will bridge this research to real-world impact by briefly showcasing how these principles are being applied at KRAFTON, powering player-assisting agents and intelligent NPCs in industrial-scale games like PUBG: BATTLEGROUNDS and inZOI. This talk establishes a clear pathway from foundational research to the deployment of sophisticated LLM agents in the wild, paving the way for the future of intelligent, interactive game experiences.

Bio: Jaewoo Ahn is a Ph.D. candidate in the Department of Computer Science and Engineering at Seoul National University where he is advised by Prof. Gunhee Kim as a member of Vision & Learning Lab. In addition, he is currently a Research Scientist Intern at KRAFTON AI. He is broadly interested in NLP and multimodal AI, with a particular focus on human-like embodied conversational agents that interact naturally in real-world environments. To this end, his work has advanced consistent persona modeling (MPChat, TimeChara), robust perception (MAC), and embodied action (Orak, FlashAdventure). Currently, his research focuses on integrating multisensory perception into LLMs to support multimodal interactions in diverse (e.g., computer-use, video game, embodied) environments. In particular, he focuses on enhancing decision-making capabilities of LLM/VLM agents across 2D/3D environments.

Program

Sunday, November 9, 2025

- 08:55 - 09:00 *Opening Remarks*
- 09:00 - 09:45 *Using generative AI to transform narrative design in games - **Sudha Rao***
- 09:45 - 10:30 *Invited Talk 2 - **Xin (Eric) Wang***
- 10:30 - 11:00 *Coffee Break (30 minutes)*
- 11:00 - 11:45 *Beyond Exams: Evaluating LLM Agents Through Interactive Collaboration and Competition - **Diyi Yang***
- 11:45 - 12:00 *Lightning Talks - Outstanding Papers*
- Wordcaster: A Gamified Speech-Based Reading Game for Young Learners Using Real-Time ASR*
Hamdan Al-Ali and Hanan Aldarmaki
- Real-Time World Crafting: Generating Structured Game Behaviors from Natural Language with Large Language Models*
Austin Drake and Hang Dong
- 12:00 - 13:15 *Lunch Break and Posters (75 minutes)*
- 13:15 - 13:45 *Lightning Talks - Spotlight Papers*
- Can They Dixit? Yes they Can! Dixit as a Playground for Multimodal Language Model Capabilities*
Nishant Balepur, Dang Nguyen and Dayeon Ki
- TextGames: Learning to Self-Play Text-Based Puzzle Games via Language Model Reasoning*
Frederikus Hudi, Genta Indra Winata, Ruochen Zhang and Alham Fikri Aji
- VLM-Dixit: Investigating Multi-Modal Abductive Reasoning and Entailment Verification with VLMs in Dixit Gameplay*
MO Yunxiang, Tianshi Zheng, Qing Zong, Jiayu Liu, Baixuan Xu, Yauwai Yim, Chunkit Chan, Jiaxin Bai and Yangqiu Song
- Explore Branches the Story didn't Narrate: An LLM Solution*
Shijun Xu, Xiangbing Kong, Danyang Chen and Zhi Liu

Sunday, November 9, 2025 (continued)

- 13:45 - 14:45 *Common-sense Persona-Grounded Dialogue Challenge (CPDC) 2025*
- Iterative Prompt Engineering for NPC Dialogue Generation: CPDC 2025 API Track*
Yukito Minari, Sei Ueno Ueno and Akinobu Lee
- Talk Less, Call Right: Enhancing Role-Play LLM Agents with Automatic Prompt Optimization and Role Prompting*
Saksorn Ruangtanusak, Pittawat Taveekitworachai and Kunat Pipatanakul
- Efficient Tool-Calling Multi-Expert NPC Agent for Commonsense Persona-Grounded Dialogue*
Mahammad Nuriyev
- Model Fusion with Multi-LoRA Inference for Tool-Enhanced Game Dialogue Agents*
Kangxu Wang, Ze Chen, Chengcheng Wei, Jiewen Zheng, Jiarong He and Max Gao
- Two-Stage Tool-Enhanced NPC Dialogue for CPDC 2025: GPU-Track Submission to Task 1 and Task 2*
Yiyang Zheng
- Interactive AI NPCs Powered by LLMs: Technical Report for the CPDC Challenge 2025*
Yitian Huang, Yuxuan Lei, Jianxun Lian and Hao Liao
- Deflanderization for Game Dialogue: Balancing Character Authenticity with Task Execution in LLM-based NPCs*
Pasin Buakhaw, Kun Kerdthaisong, Phuree Phenhiran, Pitikorn Khlaisamniang, Supasate Vorathammathorn, Nuchanon Yongsatianchot and Piyalitt Ittichaiwong
- 14:45 - 15:30 *Invited Talk 4 - **Boyang Li***
- 15:30 - 16:00 *Coffee Break (30 minutes)*
- 16:00 - 16:45 *LLM Agents for In-the-Wild Video Game-Playing - **Jaewoo Ahn***
- 16:45 - 17:00 *Closing Remarks*

Author Index