



SIGGRAPH
ASIA 2024
TOKYO

東京

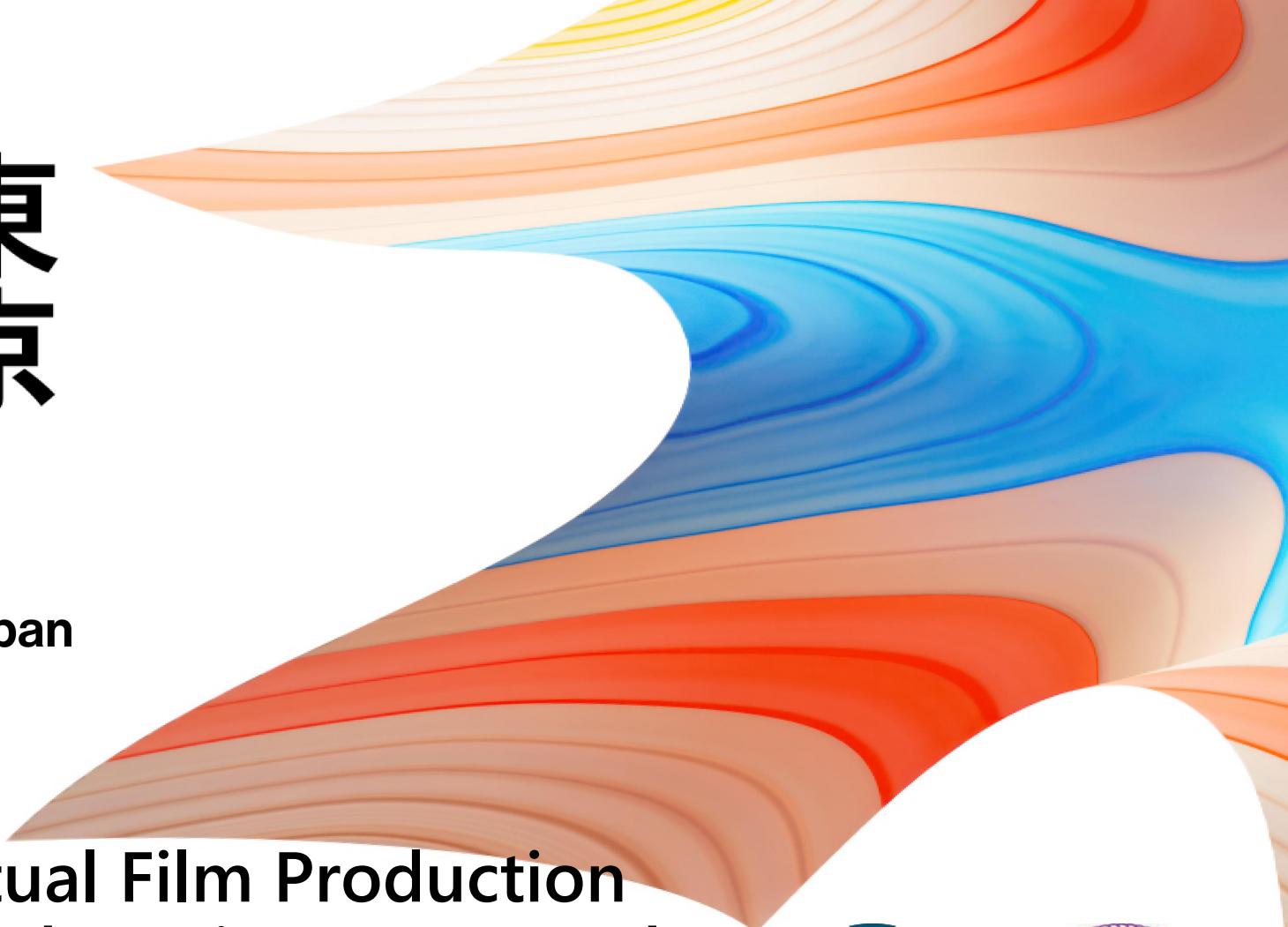
Conference | 3–6 December 2024

Exhibition | 4–6 December 2024

Venue | Tokyo International Forum, Japan

FilmAgent: Automating Virtual Film Production Through a Multi-Agent Collaborative Framework

Zhenran Xu, Jifang Wang, Longyue Wang,
Zhouyi Li, Senbao Shi, Baotian Hu, Min Zhang

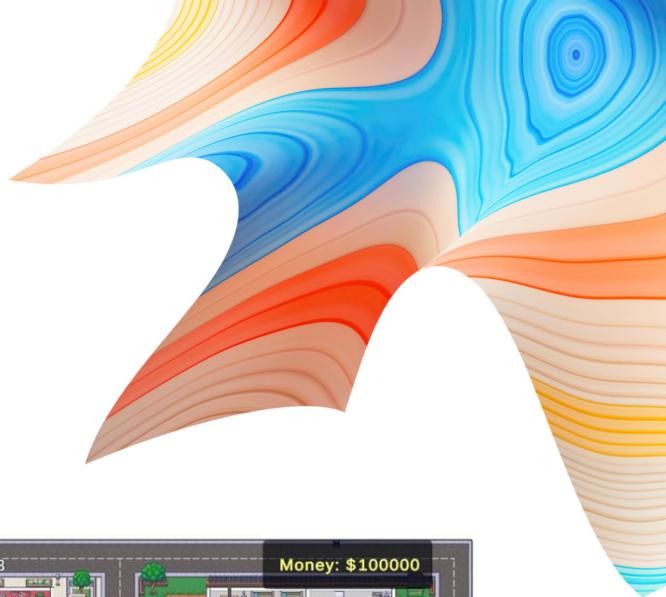


Sponsored by



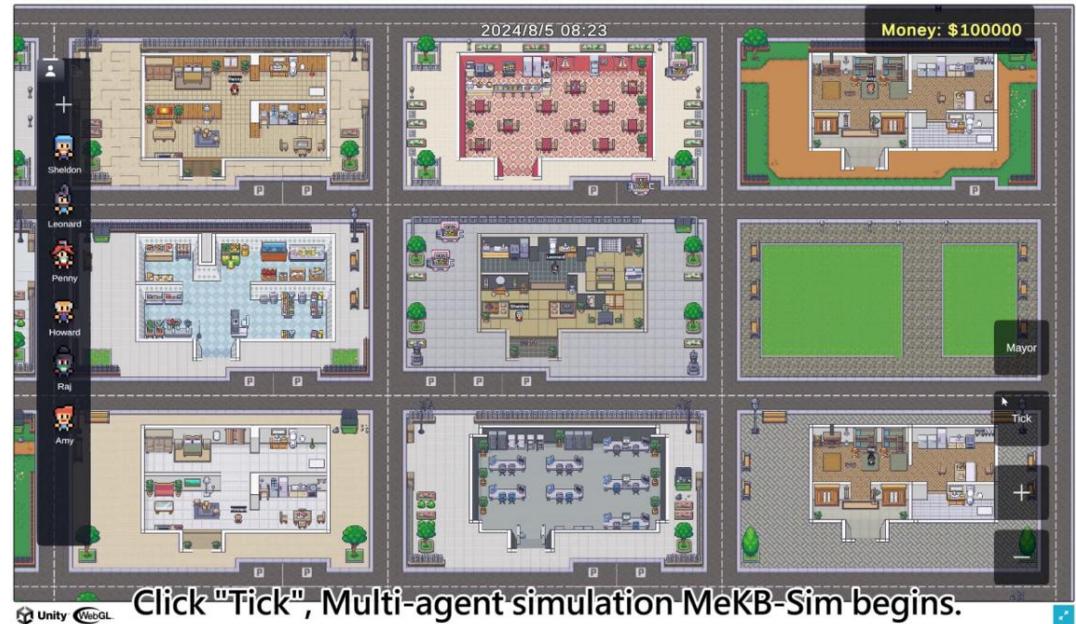
Organized by





Motivation

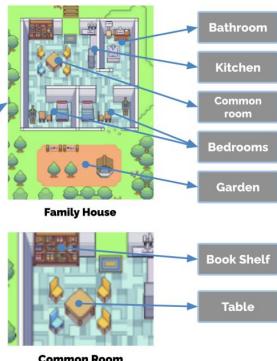
- Starting from “Stanford Town”:
 - Agents are making their daily plans and autonomously moving in the map.
 - Multiple emergent social behaviors are found: throwing gatherings, information diffusion, etc.



Park J S, O'Brien J, Cai C J, et al. Generative agents: Interactive simulacra of human behavior. 2023.

Motivation

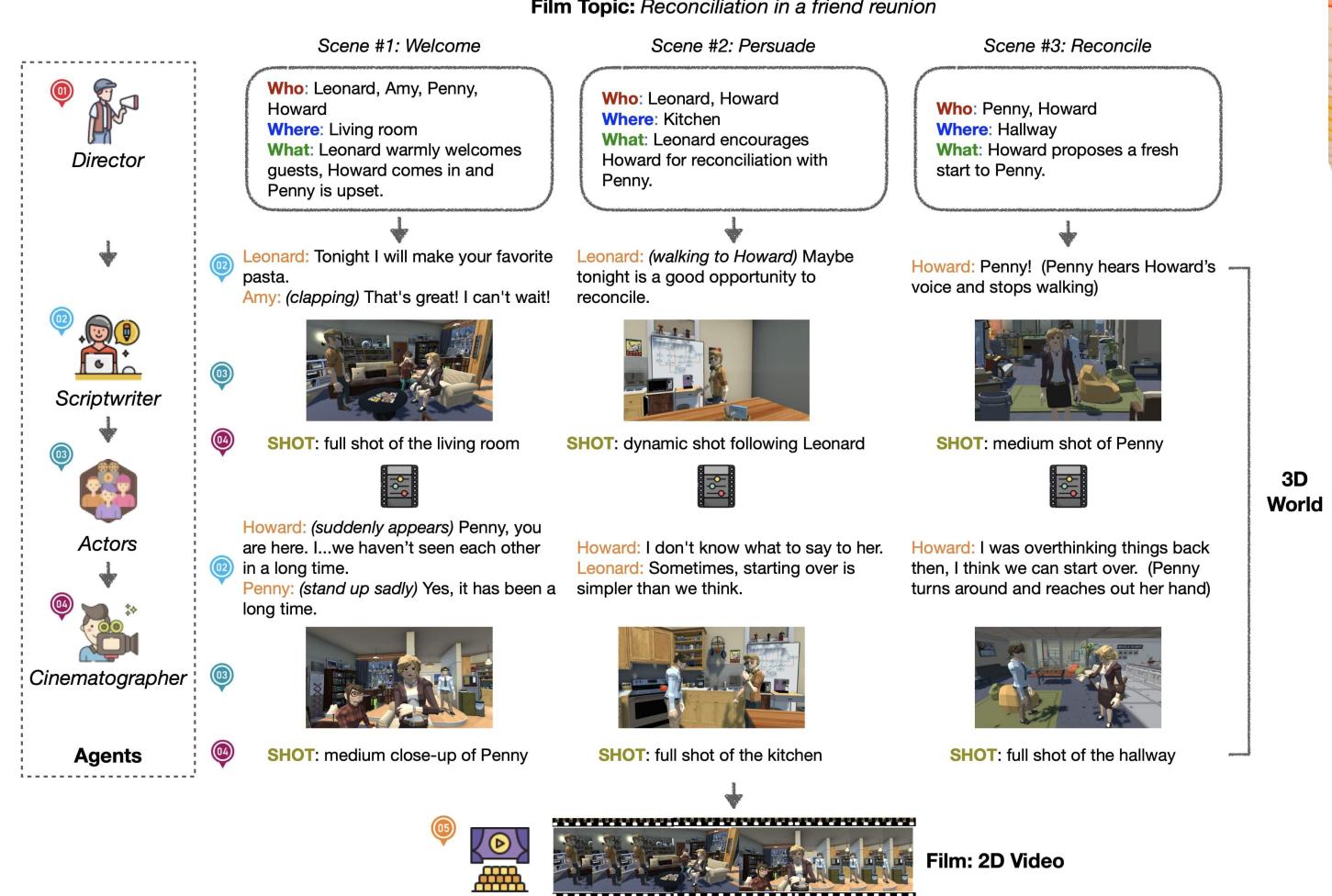
- What if we let the agents work together for shooting a video?
 - Current end-to-end text-to-video models cannot generate long videos with a detailed plot.
 - Virtual Film Production is a **communication-driven collaborative task** involving directing, camera placement and actor positioning.



From “Simulated Town” to “FilmAgent”:
 2D → 3D; Simulation without intervention → Working towards a goal

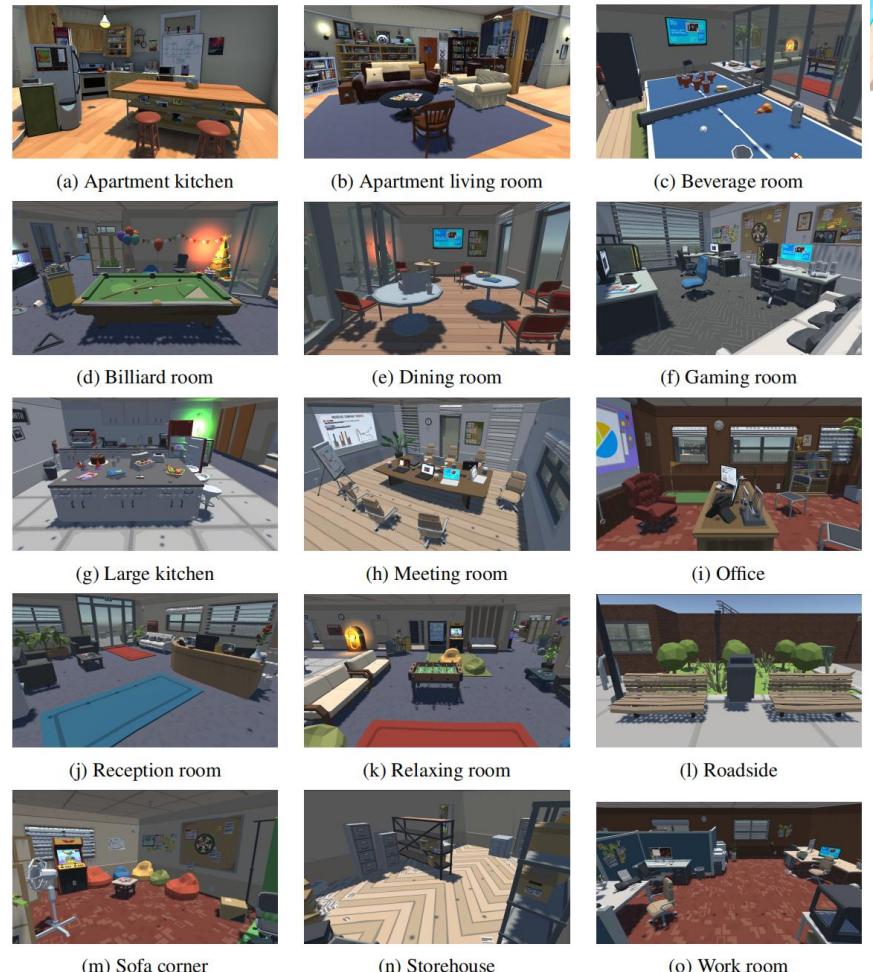
FilmAgent: Overview

- Simulate important LLM-based crew roles
 - Director
 - Scriptwriter
 - Actors
 - Cinematographer
- Each agent is specified with their own roles, duties and specific knowledge.



FilmAgent: “Film Set” Construction

- Scenes: 15 locations that reflect everyday settings





FilmAgent: “Film Set” Construction

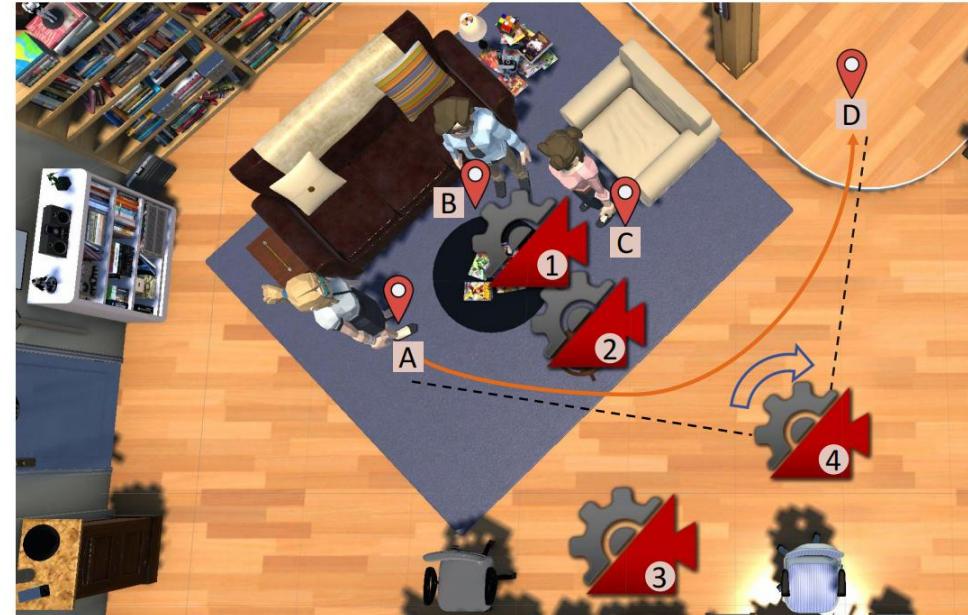
- Scenes: 15 locations that reflect everyday settings
- Actions: 21 actions from basic movements to expressive gestures

Action Space (21 actions)



FilmAgent: “Film Set” Construction

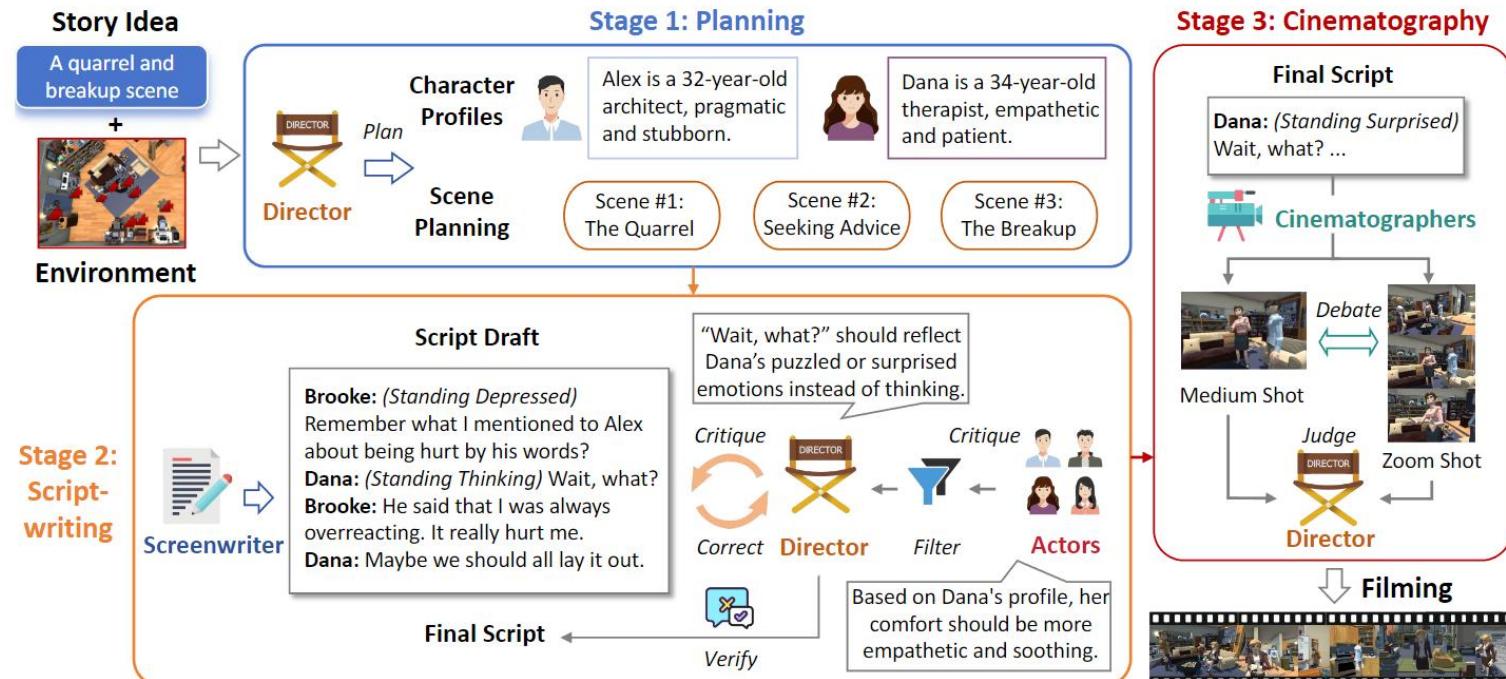
- **Scenes:** 15 locations that reflect everyday settings
- **Actions:** 21 actions from basic movements to expressive gestures
- **Positions:** 32 standing points and 33 sitting points
- **Cameras:** 165 static shots: Close-up, medium, long shots; 107 dynamic shots: 9 types (including pan, zoom, tracking shots)



- Actor Positions
- Actor Movement
- Cameras
 - 1 Close-up Shot of Position B
 - 2 Medium Shot of Position B
 - 3 Full Shot of Position B
 - 4 Pan Shot following Actor Movement (from A to D)
- Camera Movement

FilmAgent: Workflow

- Simulate human workflows:
 - Planning: Develop ideas into outlines (who, what, where)
 - Scriptwriting: Write lines for each characters, add actions, positions and possible movements
 - Cinematography: Add appropriate shots to each line



FilmAgent: Multi-Agent Collaboration

- Two multi-agent collaboration methods: Debate-Judge & Critique-Correct-Verify
- The multi-agent feedback and revision have improved the plot coherence, action accuracy and camera appropriateness.

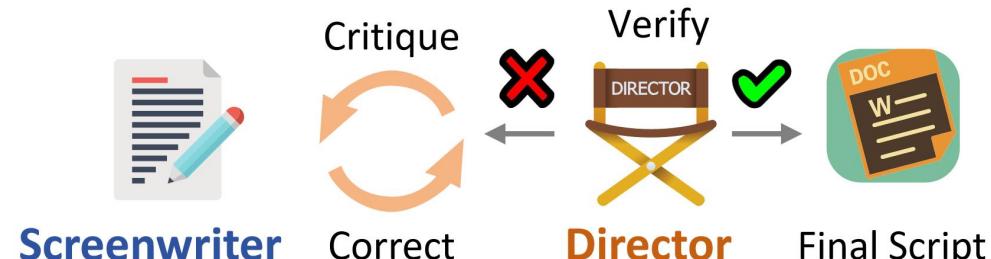


Cinematographers



Director

Debate-Judge



Screenwriter

Critique-Correct-Verify

FilmAgent: Multi-Agent Collaboration

- Two multi-agent collaboration methods: Debate-Judge & Critique-Correct-Verify
- The multi-agent feedback and revision have improved the plot coherence, action accuracy and camera appropriateness.

Scene #1 (Roadside)

Emma: I'd love that. Where should we meet?

Alex: (Standing suggest) There's a cafe just around the corner from here. How about tomorrow at 3?

Emma: (Standing happy) Perfect! See you tomorrow.

Scene #2 (In Alex's house)

Alex: (Sit down) Welcome to our humble abode! Make yourself comfortable.

Scene #1 (Roadside)

Emma: I'd love that. Where should we meet?

Alex: (Standing thinking) How about at my place? Tomorrow at 3?

Emma: (Standing happy) Perfect! See you tomorrow.

Scene #2 (In Alex's house)

Alex: (Sit down) Welcome to our humble abode! Make yourself comfortable.

Fixed the non-existent action “stand suggest” and adjusted the response to be consistent with Alex's house in the next scene.



Select Follow Shot, the speaker faces away from the camera



Select Mid Shot, the speaker faces the camera and body movements are reflected



Demo



3-6 December 2024

Tokyo International Forum, Japan

ASIA.SIGGRAPH.ORG/2024

Sponsored by



Organized by



Evaluation

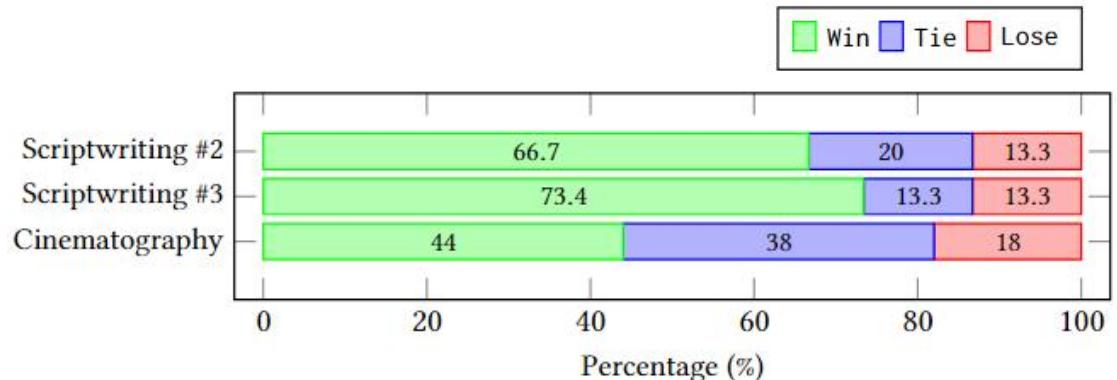
- FilmAgent surpasses all single-agent baselines across 4 aspects:
 - the appropriateness of **camera settings**
 - the alignment of the script with **actor profiles**
 - the accuracy of **actor actions**
 - the overall **plot coherence**.

	Action	Plot	Profile	Camera
CoT	0.62	1.33	3.26	1.67
FilmAgent (Solo)	0.80	1.87	4.20	2.07
FilmAgent (Group)	0.88	3.53	4.44	3.53

Evaluation

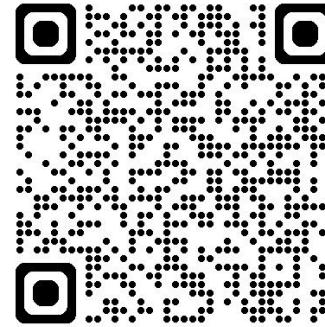
- FilmAgent surpasses all single-agent baselines across 4 aspects:
 - the appropriateness of **camera settings**
 - the alignment of the script with **actor profiles**
 - the accuracy of **actor actions**
 - the overall **plot coherence**.
- Human preference evaluation shows a clear preference for the revised scripts over the original versions.

	Action	Plot	Profile	Camera
CoT	0.62	1.33	3.26	1.67
FilmAgent (Solo)	0.80	1.87	4.20	2.07
FilmAgent (Group)	0.88	3.53	4.44	3.53

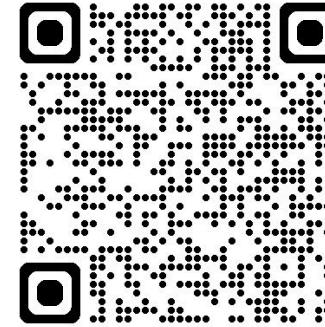


Limitations & Future Work

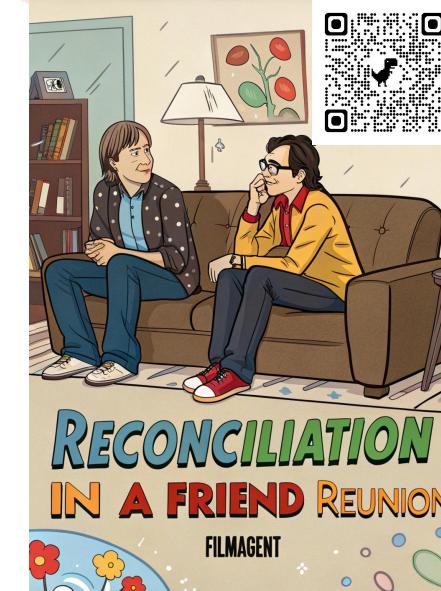
- The biggest limitation is the **manual film set construction part**, offering too much constraints.
 - Automatic motion synthesis, camera adjustments based on language
- Video creation is inherently a **multimodal** task.
 - Let large multimodal models provide more accurate feedback and revision.
- More roles to be added: art director, colorist, lighting, music supervisor, etc.



Project Website 

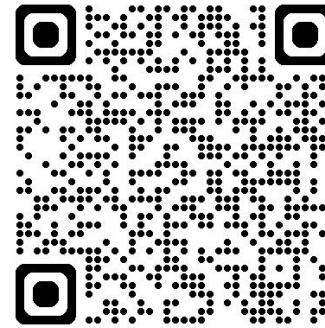


Open-sourced Toolkits 

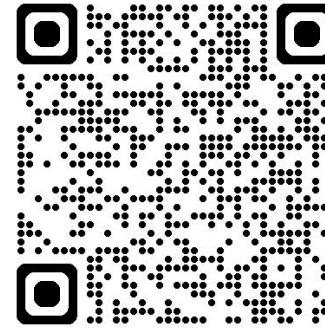


Conclusion & Take-away Message

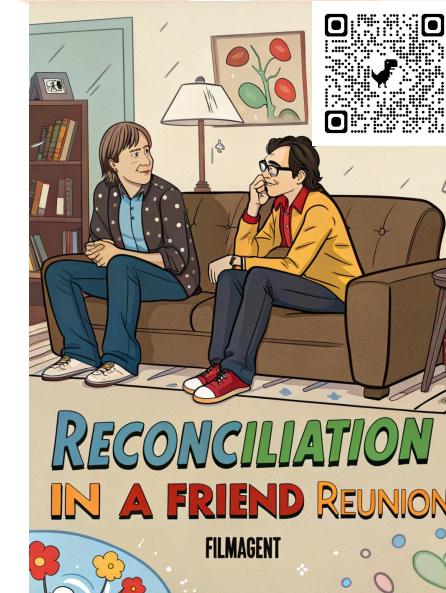
- **FilmAgent** is a preliminary tryout of LLM-based multi-agent collaboration on creative content generation, mirroring the traditional film set process.
- **Multi-agent collaboration** can help reduce hallucinations and enhances the accuracy of decision making.
- Large Language/Multimodal models act as a **planner** and bring **world knowledge**, showing great potential in such creative domain.



Project Website 



Open-sourced Toolkits 



Sponsored by



Organized by





Thank you!