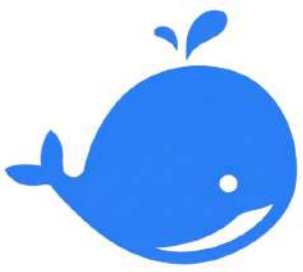
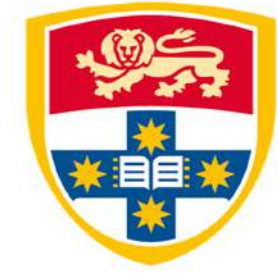
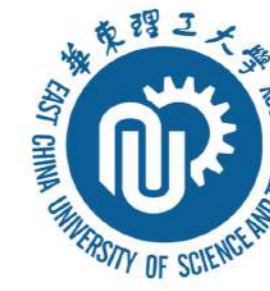




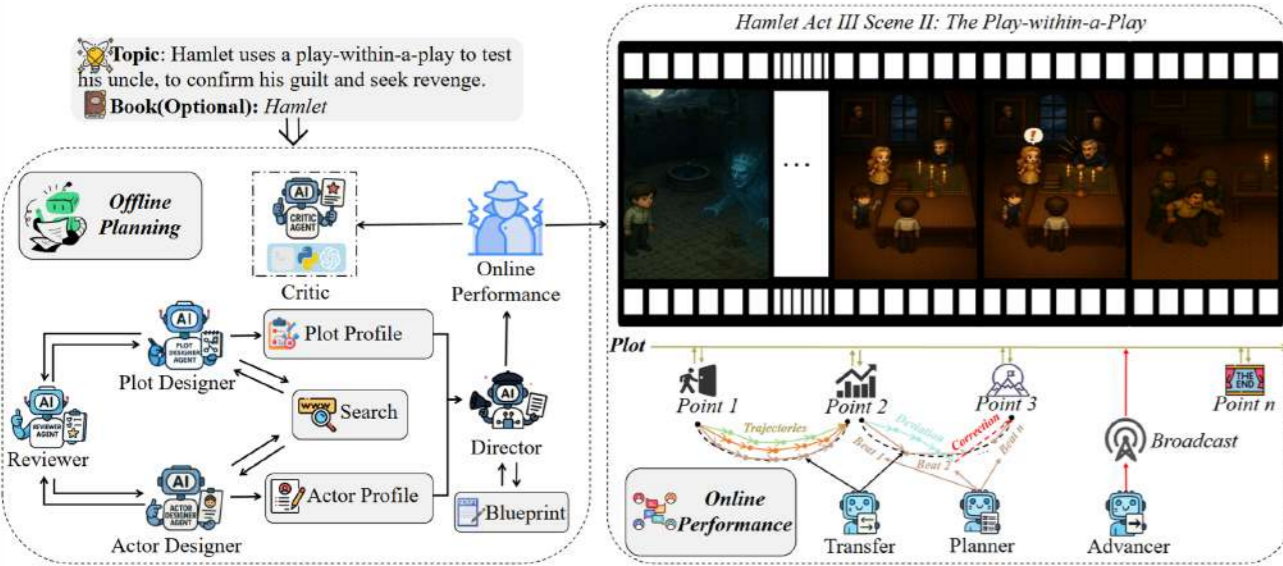
HAMLET: A Hierarchical and Adaptive Multi-Agent Framework for Live Embodied Theatrics

Shufan Jiang, Sizhou Chen, Chi Zhang, Xiao-Lei Zhang, Xuelong Li

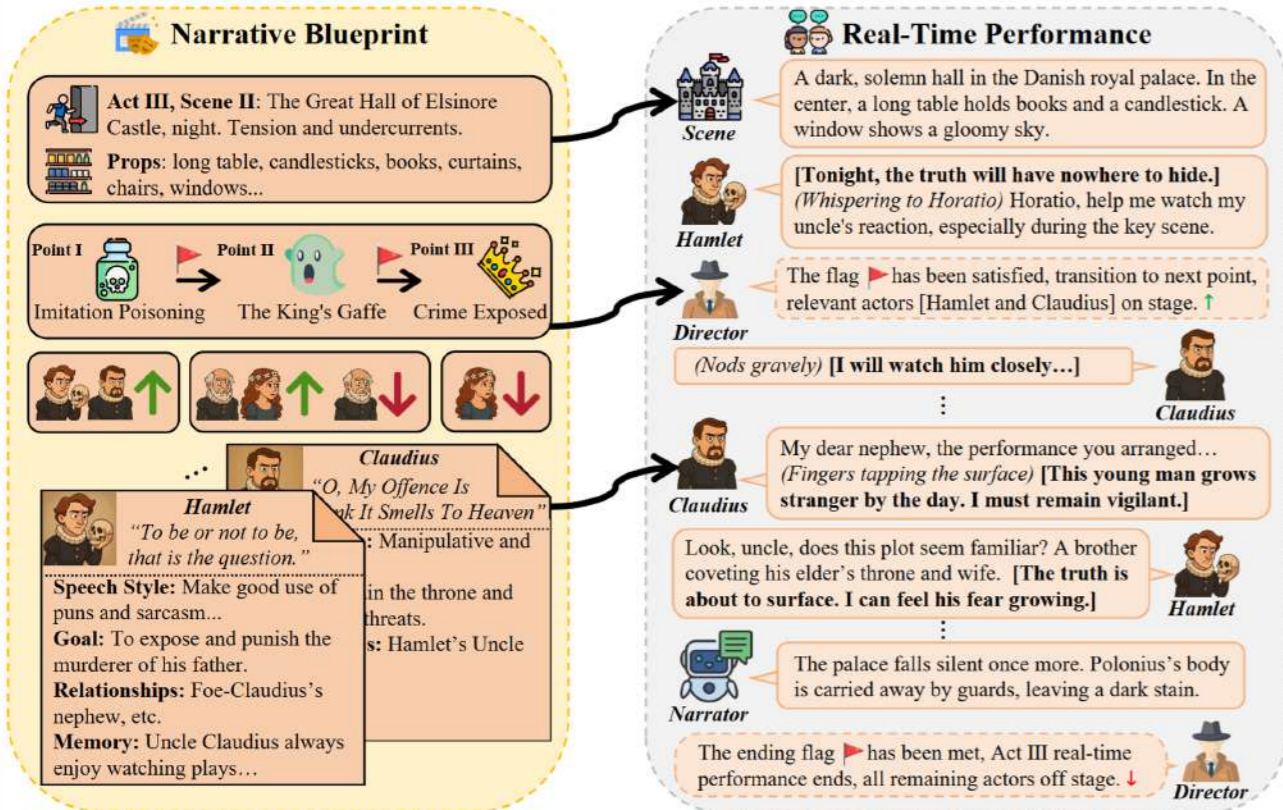


Overview

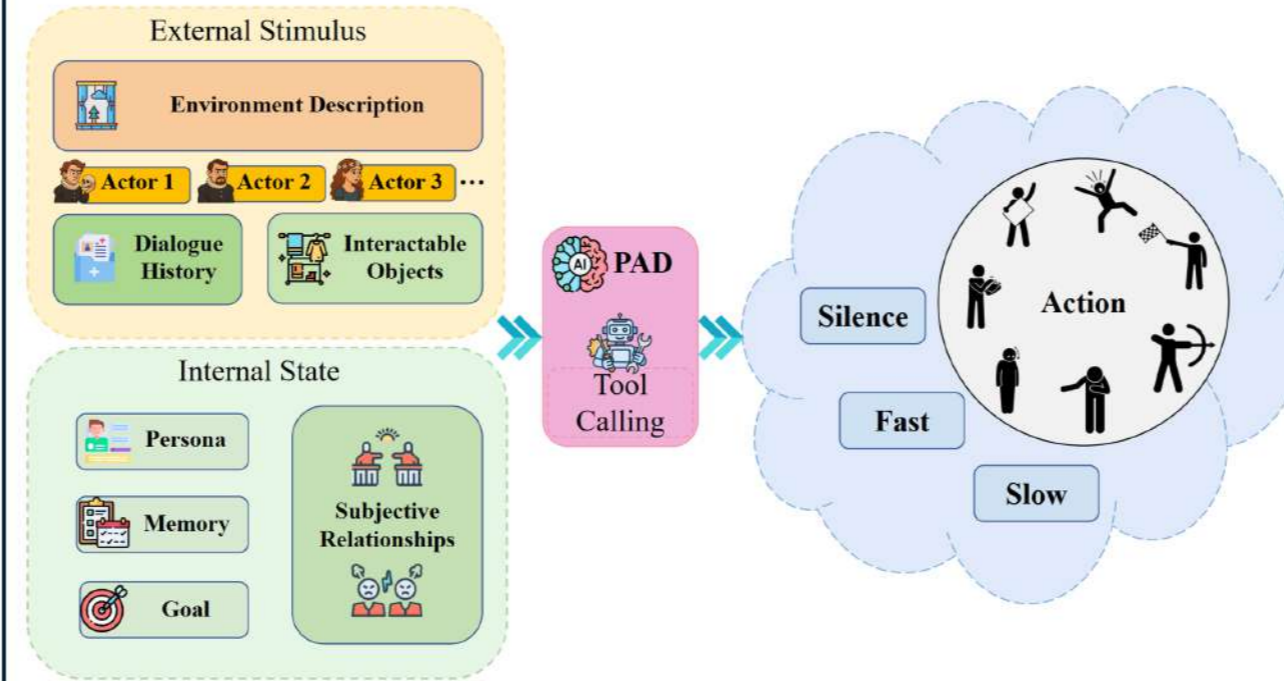
We proposed a multi-agent framework **HAMLET** that decouples **offline planning** and **online performance** in AI theatrics scenarios.



HAMLET excels in creating expressive, coherent, real-time and physically interactive drama experiences in a fully autonomous manner.

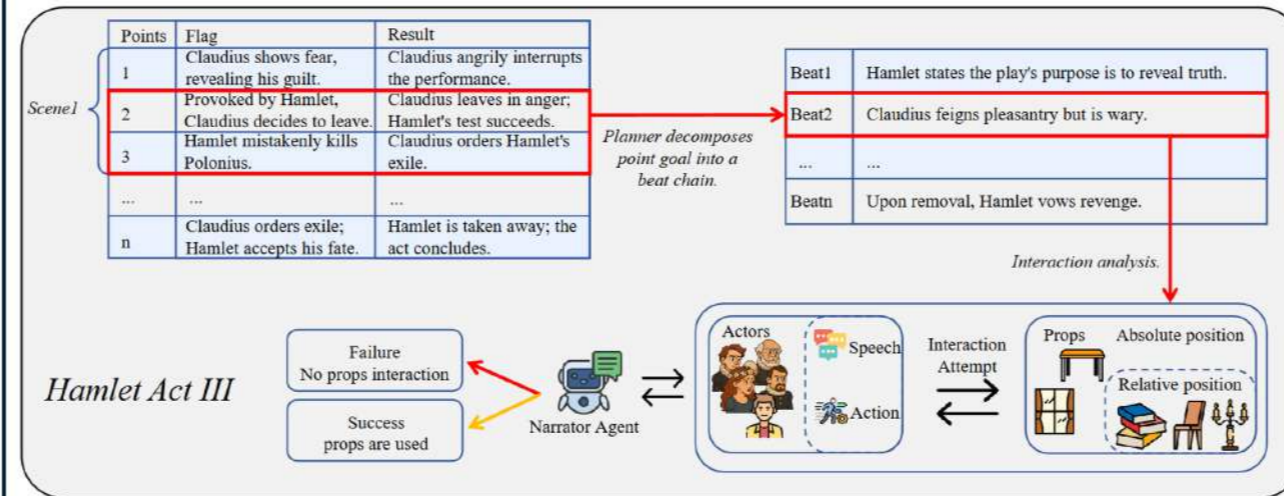


Method



Perceive And Decide Module (PAD) enables actor agents to

- Proactively engage into the session;
- Make reasonable action and control the paces of response.



Planner Agent decomposes point goal into beats
Narrator Agent ensures the rationality of physical actions
Together, they empower AI drama that --

- Allows arbitrary and customizable trajectories;
- Supports environmental interaction and change the scene.

Experiments

HAMLET leaderboard (main eval results on various LLMs)

Model	English				Chinese				Overall Score
	Character	Narrative	Interaction	Average	Character	Narrative	Interaction	Average	
<i>Non-Reasoning Models</i>									
Claude-4-sonnet	76.50	77.30	76.96	76.92	80.18	79.20	79.66	79.68	78.30
Claude-3.7-sonnet	65.80	66.94	65.98	66.24	76.00	75.20	75.48	75.56	70.90
Claude-3.5-sonnet	61.00	62.15	62.10	61.75	60.50	59.00	59.99	59.83	60.79
Gemini-2.5-flash	46.00	47.10	46.70	46.60	52.00	52.90	52.15	52.35	49.48
GPT-4.5-preview	59.21	60.00	59.92	59.71	61.50	62.50	61.91	61.97	60.84
GPT-4.1	63.50	62.49	62.98	62.99	62.00	62.80	62.64	62.48	62.74
Mistral-Small-3.2-24B	48.00	48.50	48.22	48.24	51.00	51.83	52.00	51.61	49.93
DeepSeek-V3-0324	54.00	55.13	55.00	54.71	64.50	64.00	64.01	64.17	59.44
Llama-3.1-70B	49.80	49.00	48.86	49.22	42.00	42.81	42.00	42.27	45.75
Llama-3.1-8B	35.00	36.01	35.52	35.51	33.50	34.00	33.99	33.83	34.67
Higgs-Llama-3-70B	72.00	78.50	66.22	72.24	64.00	64.50	63.95	64.15	68.20
Qwen3-8B	47.50	46.80	47.00	47.10	58.00	57.50	58.05	57.85	52.48
Qwen3-32B	65.00	65.88	65.20	65.36	66.00	65.50	65.84	65.78	65.57
Qwen3-235B-A22B	69.50	69.80	69.65	69.65	72.50	73.00	72.78	72.76	71.21
<i>Reasoning Models</i>									
Gemini-2.5-pro	61.00	62.22	61.70	61.64	62.00	62.80	62.25	62.35	62.00
Claude-4-sonnet-Thinking	79.50	78.40	79.04	78.42	80.32	81.03	79.92	79.45	79.45
Minimax-M1	51.50	52.32	52.00	51.94	65.00	65.19	65.23	65.23	58.59
OpenAI-o3	69.00	69.95	69.40	69.45	78.00	77.50	78.17	77.89	73.67
Magistral-Small-2506	59.00	60.00	59.74	59.58	58.50	59.30	58.90	58.90	59.24
DeepSeek-R1-0528	66.00	67.10	66.64	66.58	79.00	79.61	79.37	79.28	72.98
Qwen3-8B-Thinking	50.00	51.61	51.00	50.87	65.00	65.55	65.45	65.45	58.16
Qwen3-32B-Thinking	69.50	68.80	69.00	69.10	78.00	79.00	78.77	78.59	73.85
Qwen3-235B-A22B-Thinking	70.50	71.00	70.72	70.74	76.00	75.80	75.96	75.92	73.33

Ablation Study

