

# Generative Agents: Interactive Simulacra of Human Behavior

arXiv:2304.03442v2 [cs.HC] 6 Aug 2023

UIST '23, October 29-November 1, 2023, San Francisco, CA, USA

-박지영

Taking a walk  
in the park

SM: ☺

Joining for coffee at a cafe

KM: ☺

AC: ☺

[Abigail]: Hey Klaus, mind if  
I join you for coffee?  
[Klaus]: Not at all, Abigail.  
How are you?

Arriving at school

AK: ☺

Sharing news with colleagues

JL: ☺

TM: ☺

[John]: Hey, have you heard  
anything new about the  
upcoming mayoral election?  
[Tom]: No, not really. Do you  
know who is running?

Finishing a  
morning routine

JH: ☺

# 목차

1 ABSTRACT

2 SANDBOX  
ENVIRONMENT  
IMPLEMENTATION

3 Generative  
Agent Architecture

4 EVALUATION1 -  
CONTROLLED

5 EVALUATION2 -  
END-TO-END

6 DISSCUSSION

7 CONCLUSION

# 1. ABSTRACT

- Generative AI of this paper

LLM + interactive agent = believable simulation architecture

- Baseline

GPT 3.5

- Difference of previous paper

최종 사용자는 에이전트들을 관찰, 상호 작용

Ex) 예를 들어, 최종 사용자나 개발자가 마을에서 게임 내 발렌타인 데이 파티를 개최하기를 원한다면, 전통적인 게임 환경에서는 수십 명의 캐릭터의 행동을 수동으로 스크립팅해야 합니다. 저희는 생성 에이전트를 사용하면 한 에이전트에게 파티를 열고 싶다고 간단히 말하는 것으로 충분하다는 것을 보여줍니다.”



## 2. SANDBOX ENVIRONMENT IMPLEMENTATION

- Framework : Phaser
- Server: sandbox



Family House



Common Room

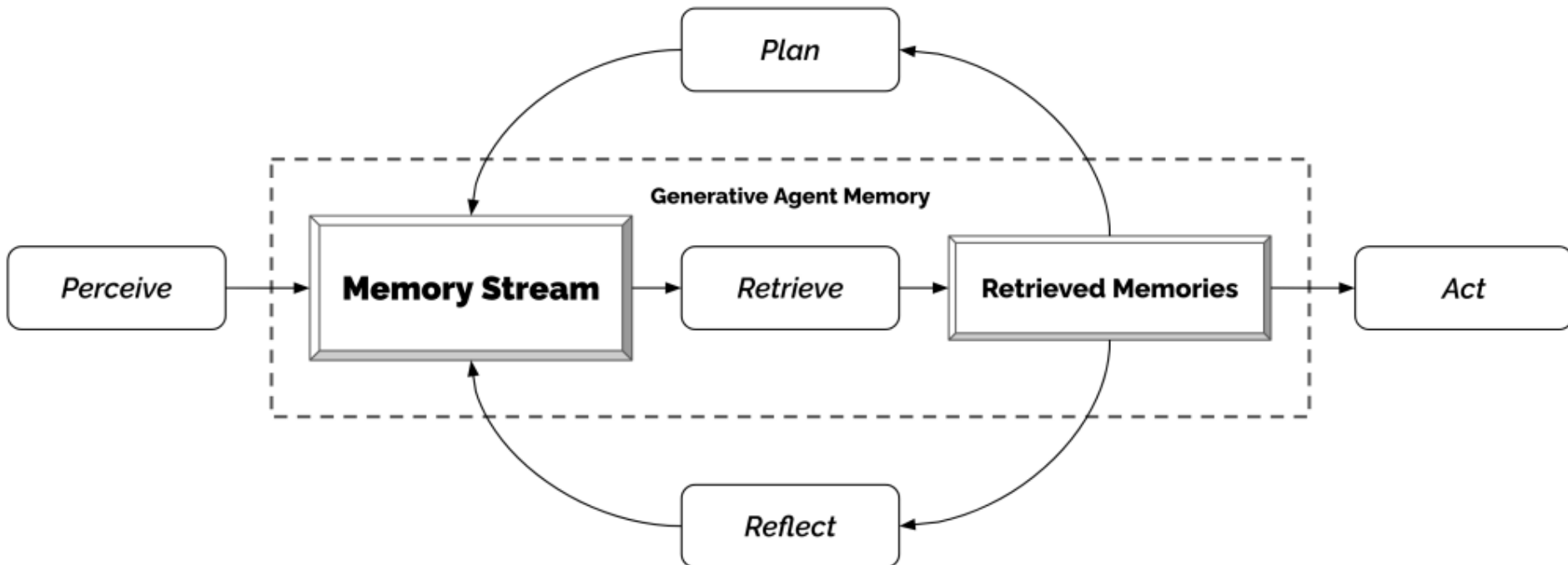
Book Shelf

Table

### 3. Generative Agent Architecture

Three main components.

- Memory stream
- Reflection
- Plan



### 3. Generative Agent Architecture - Memory stream

### Memory Stream

2023-02-13 22:48:20: desk is idle  
2023-02-13 22:48:20: bed is idle  
2023-02-13 22:48:10: closet is idle  
2023-02-13 22:48:10: refrigerator is idle  
2023-02-13 22:48:10: Isabella Rodriguez is stretching  
2023-02-13 22:33:30: shelf is idle  
2023-02-13 22:33:30: desk is neat and organized  
2023-02-13 22:33:10: Isabella Rodriguez is writing in her journal  
2023-02-13 22:18:10: desk is idle  
2023-02-13 22:18:10: Isabella Rodriguez is taking a break  
2023-02-13 21:49:00: bed is idle  
2023-02-13 21:48:50: Isabella Rodriguez is cleaning up the kitchen  
2023-02-13 21:48:50: refrigerator is idle  
2023-02-13 21:48:50: bed is being used  
2023-02-13 21:48:10: shelf is idle  
2023-02-13 21:48:10: Isabella Rodriguez is watching a movie  
2023-02-13 21:19:10: shelf is organized and tidy  
2023-02-13 21:18:10: desk is idle  
2023-02-13 21:18:10: Isabella Rodriguez is reading a book  
2023-02-13 21:03:40: bed is idle  
2023-02-13 21:03:30: refrigerator is idle  
2023-02-13 21:03:30: desk is in use with a laptop and some papers on it  
...

**Q. What are you looking forward to the most right now?**

Isabella Rodriguez is excited to be planning a Valentine's Day party at Hobbs Cafe on February 14th from 5pm and is eager to invite everyone to attend the party.

retrieval		recency		importance		relevance
2.34	=	0.91	+	0.63	+	0.80

ordering decorations for the party

2.21	=	0.87	+	0.63	+	0.71
------	---	------	---	------	---	------

researching ideas for the party

2.20	=	0.85	+	0.73	+	0.62
------	---	------	---	------	---	------

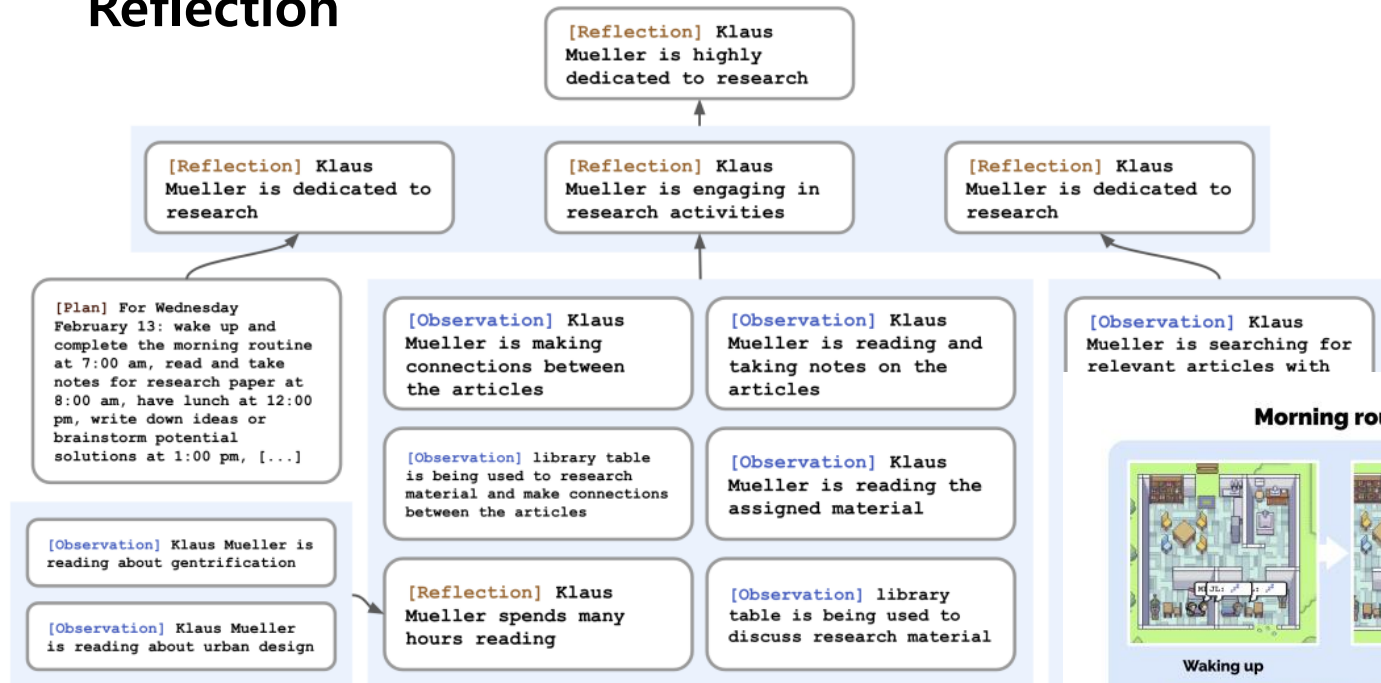
...

I'm looking forward to the Valentine's Day party that I'm planning at Hobbs Cafe!

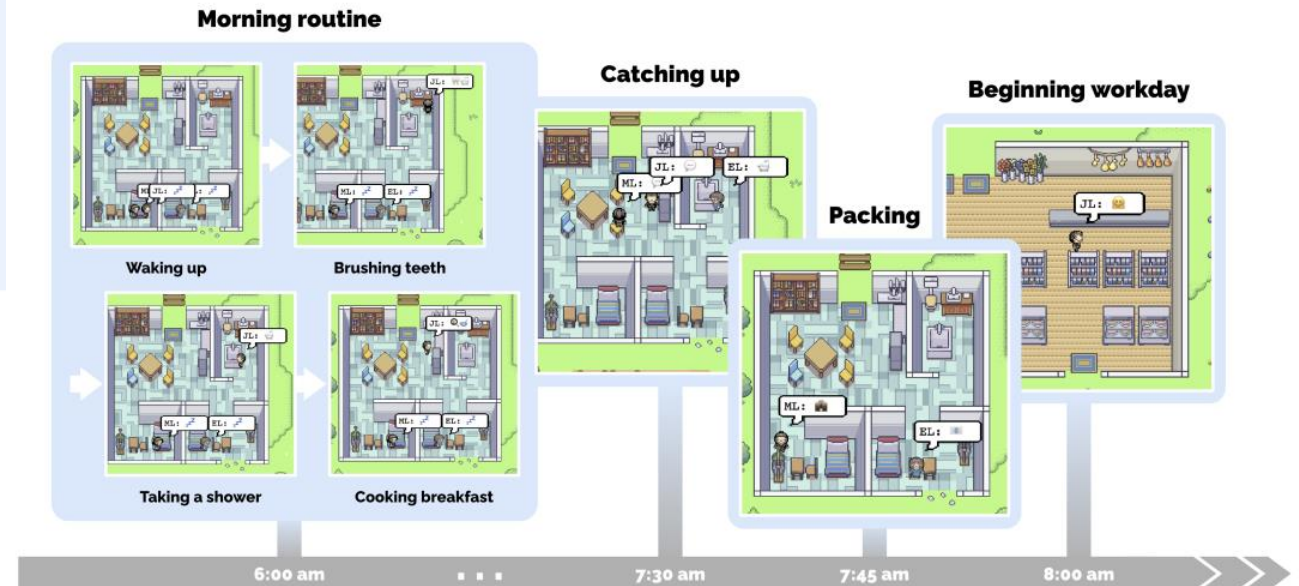


# 3. Generative Agent Architecture – reflection, plan

## Reflection



## Plan

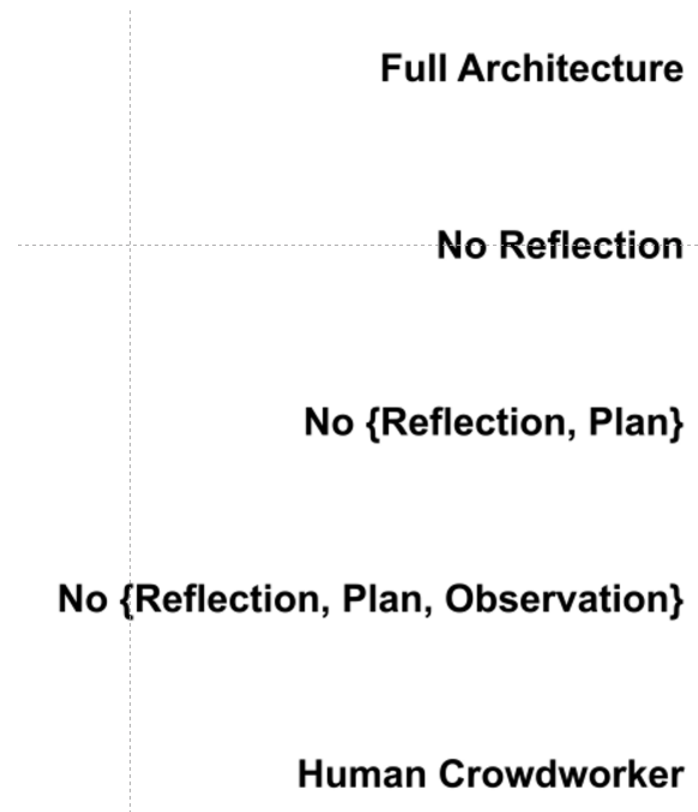


## 4. EVALUATION 1 - CONTROLLED

- 평가 기준이 되는 질문 카테고리 5개

self-knowledge, retrieving memory, generating plans, reacting, and reflecting.  
기반으로 질문을 하여 응답 생성.

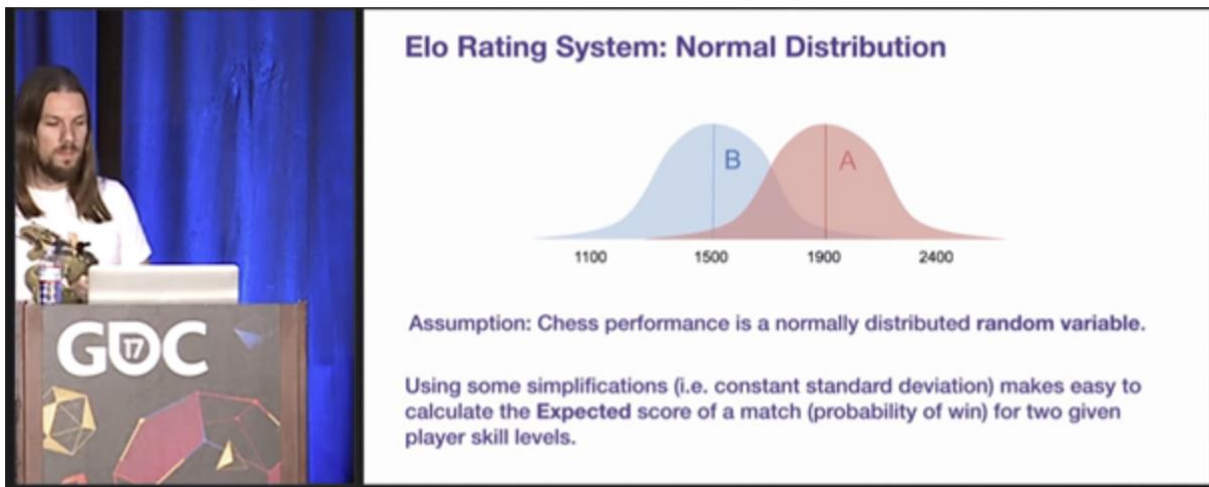
- 5가지 agents들을 평가하여 ranked  
>> trueskill 사용





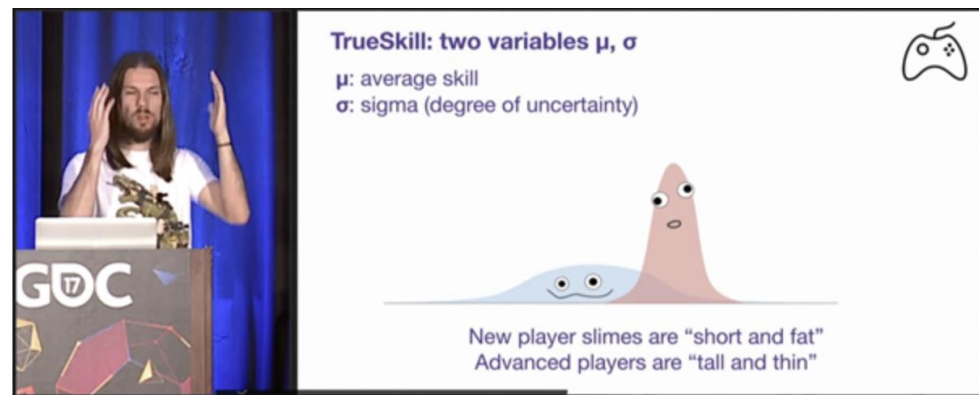
# 4. EVALUATION 1 – ELO, TrueSkill

## 기본 개념

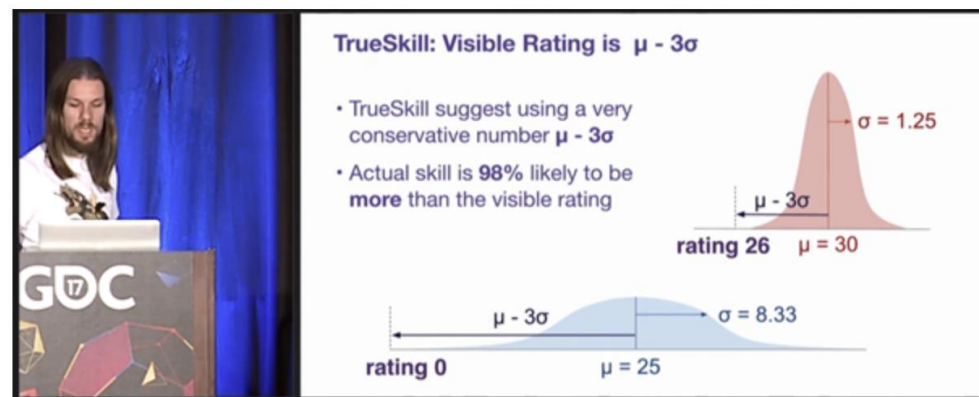


ELO는 개별 플레이어 실력의 평균( $\mu$ ) 및 표준편차( $\sigma$ ) 값을 정규분포한다고 가정한다.

- ELO는, 플레이어의 체스 실력이 **정규 분포**<sup>2</sup>한다고 가정하였다.
- 이와 같은 가정으로 인해, "플레이어의 실력이 위치할 수 있는 확률 범위"에 대한 계산이 쉬워졌고, 2명 간 매치 시의 승률 계산이 단순화될 수 있었다.
- 한편, 체스를 기반으로 한 레이팅 공식이므로, 1vs1 상황만을 가정한 공식이다.



플레이어의 실력은 정규 분포를 따르지 않는다!



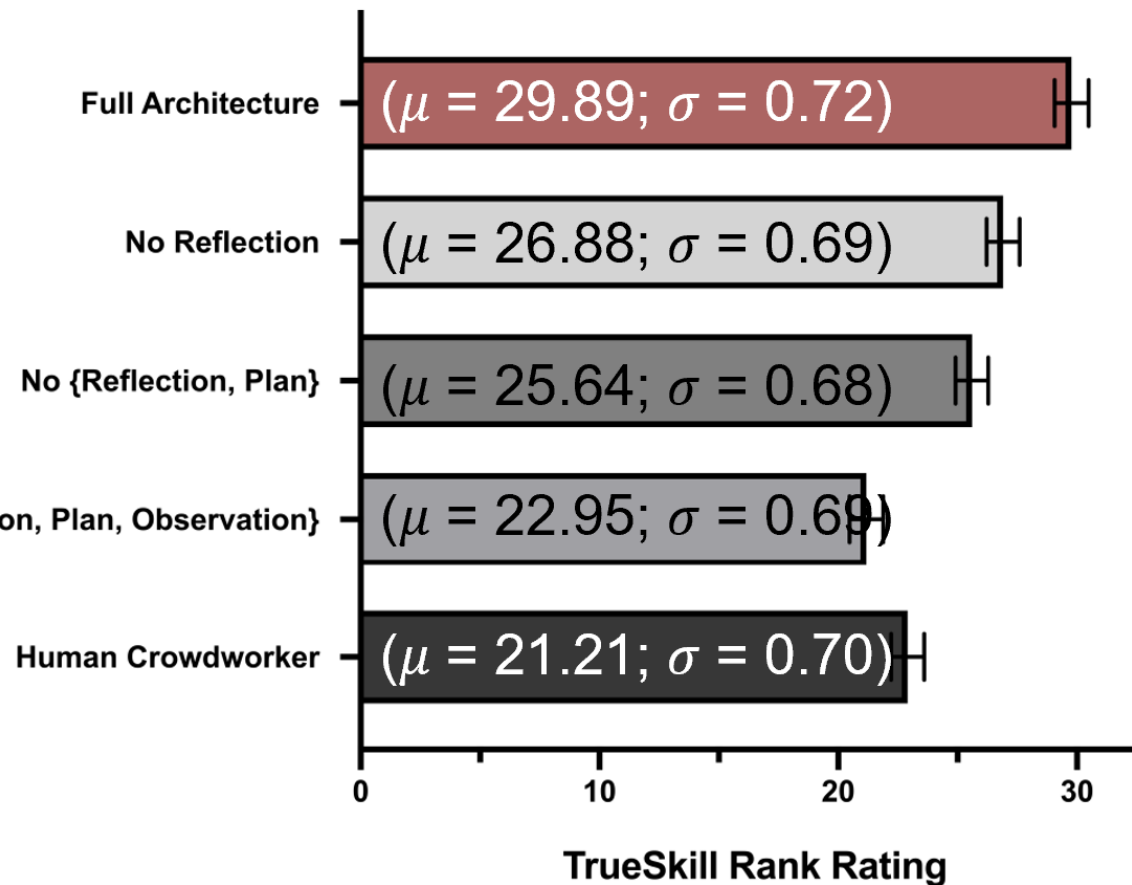
범위 값은 98% 신뢰도를 가지도록 평균( $\mu$ ) 및 표준편차( $\sigma$ ) 설정 필요

## 4. EVALUATION 1 CONTROLLED

---

RESULT:

Reflection이  
행동에 대한 결정을 내릴 때,  
생성 에이전트에게 중요한 요소



## 4. EVALUATION 1 – statistical test

*To investigate the statistical significance of these results,*

비모수 검정: 데이터가 정규 분포 따르지 않을 때 사용되며,

일원 분산 분석: 그룹 간의 평균 차이가 통계적으로 유의미?

### 1. Kruskal-Wallis 검정 >> ( $H = 150.29, p < 0.001$ )

다섯 가지 조건 간의 순위 차이의 통계적 유의성

$H_0$ (가설): 조건 간 순위 차이의 전반적인 차이가 있다.

### 2. Dunn 사후 검정 >> ( $p < 0.001$ )

X vs. Y

Y vs. Z

### 3. Holm-Bonferroni 방법 >> ( $p < 0.001$ )

X vs. Y vs. Z

→ 즉, 앞의 자료들이 타당성을 가진다.

# 5. EVALUATION 2 - END-TO-END

## End-to-end:

사용자의 개입 없이 이루어짐

## 확인할 부분:

정보 확산 + 새로운 관계 형성

## 방법:

무방향 그래프

Vertex,  $V$ : 에이전트.

Edge,  $E$ : 상호 지식

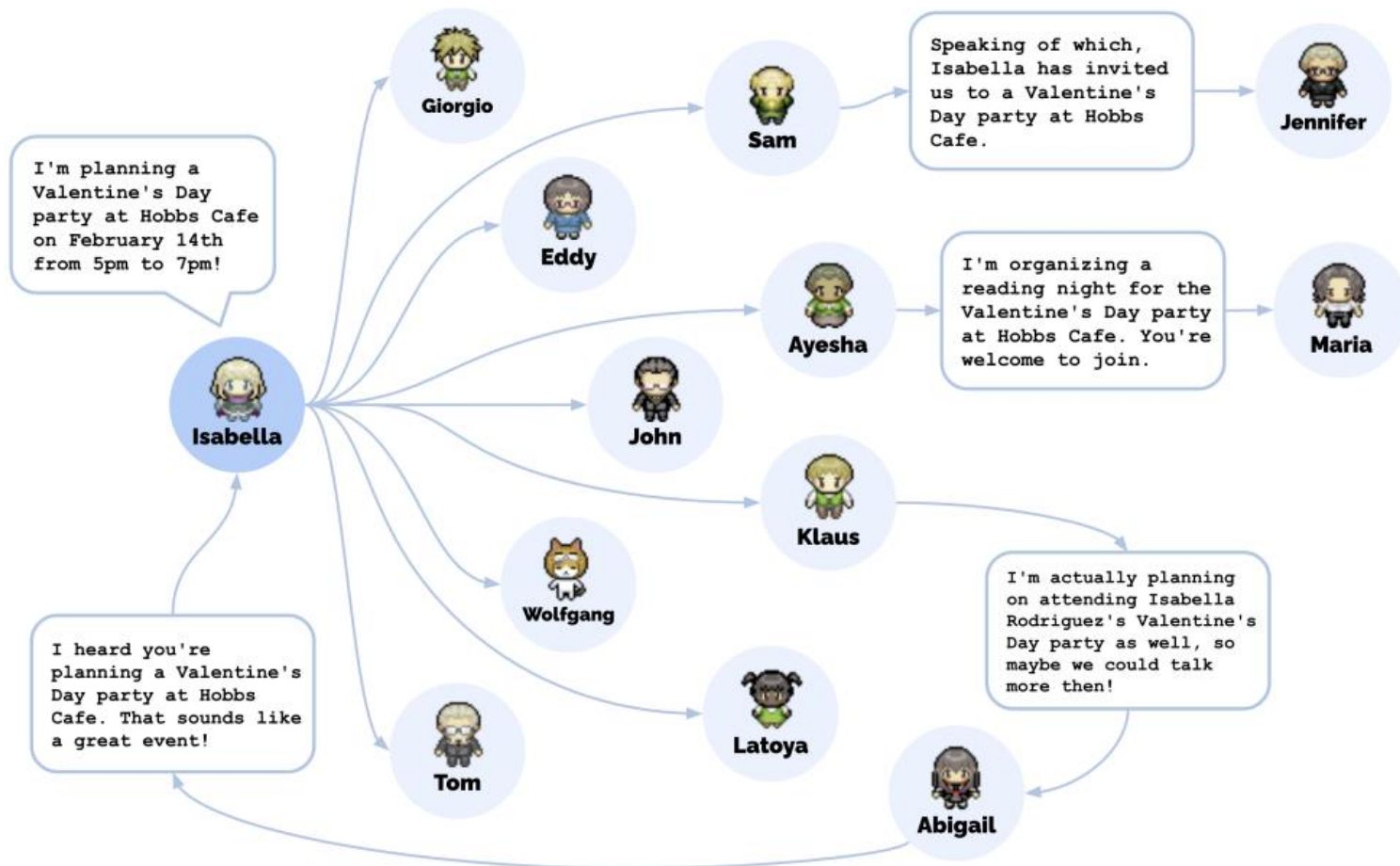
네트워크 밀도:

$$\eta = 2 * |E| / |V|(|V| - 1)$$

## 결과:

정보 확산 4% → 32%~52%

네트워크 밀도가 0.167에서 0.74로 증가



# 6. DISSCUSTION

## 1. Limitations

- 평가 측면에서 연구 추가 진행 필요

## 2. Ethics and Societal Impact

- people forming parasocial relationships with generative agents
- impact of errors
- over-reliance
- deepfakes, misinformation generation, and tailored persuasion



# 7. CONCLUSION

상호작용 컴퓨터 에이전트 소개  
경험 기록, reflection, 환경 이해를  
바탕으로 생성 에이전트 구성  
아키텍처 신뢰할 수 있는 행동 생성

더 개인화, 효과적인 기술 경험  
인지 모델 : 인간 중심 설계 프로세스  
Ex) GOMS, klms  
다양한 상호작용 응용 분야에서 역할



THANK YOU