



**KAIST FALL 2025**

# **CS473: INTRO TO SOCIAL COMPUTING**

**SOCIAL.CSTLAB.ORG**

## Lecture 19: Simulations (part 2)

**2025.12.04**

**Joseph Seering**

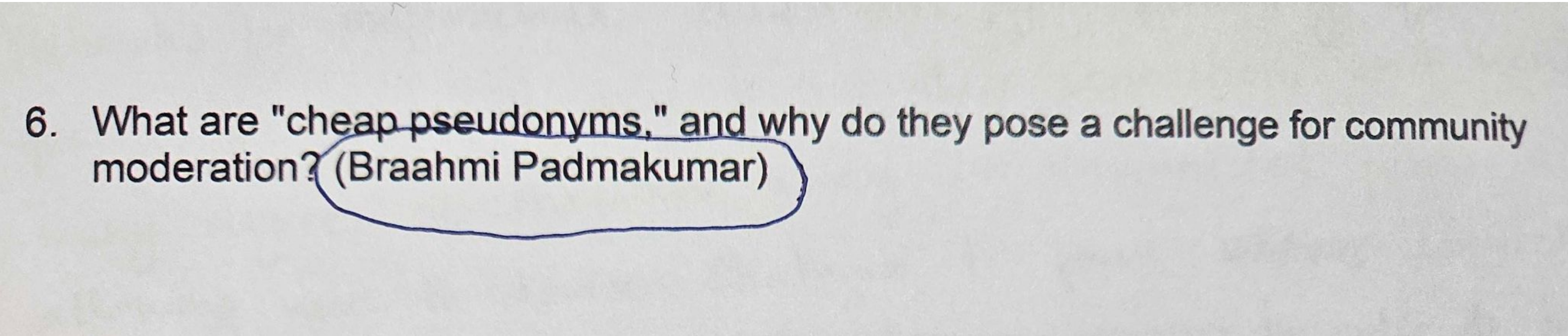
# PREVIEW OF DPM5: FINAL PRESENTATIONS

DPM5 studio (Final Presentations) on December 11th in this room!

- 5 minutes to present the problem, the core tasks you're trying to support, how your solution works and how it's better than alternatives, and how your deployment went.
- There is no Q&A during presentations.
- After presentations, there will be a 30 minute demo session where everyone in the class will try what you built.

# FINAL EXAM

- Next Tuesday, the exam question bank will be released.
- The exam will be entirely made of (slightly modified) questions you submitted.



6. What are "cheap pseudonyms," and why do they pose a challenge for community moderation? (Braahmi Padmakumar)

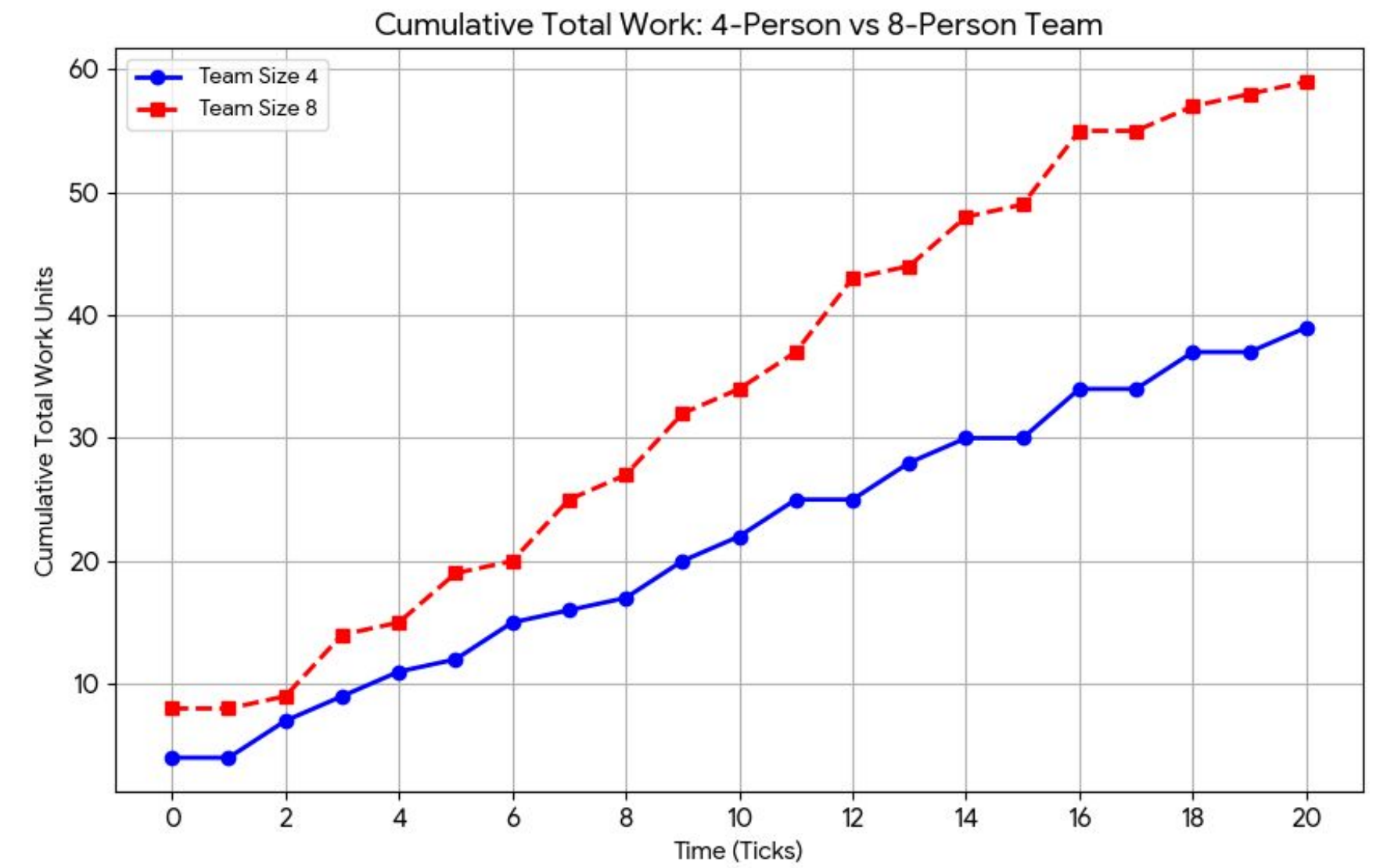
# Activity: Make a simple agent-based simulation

<https://bit.ly/CS473GroupProjectSimulation>



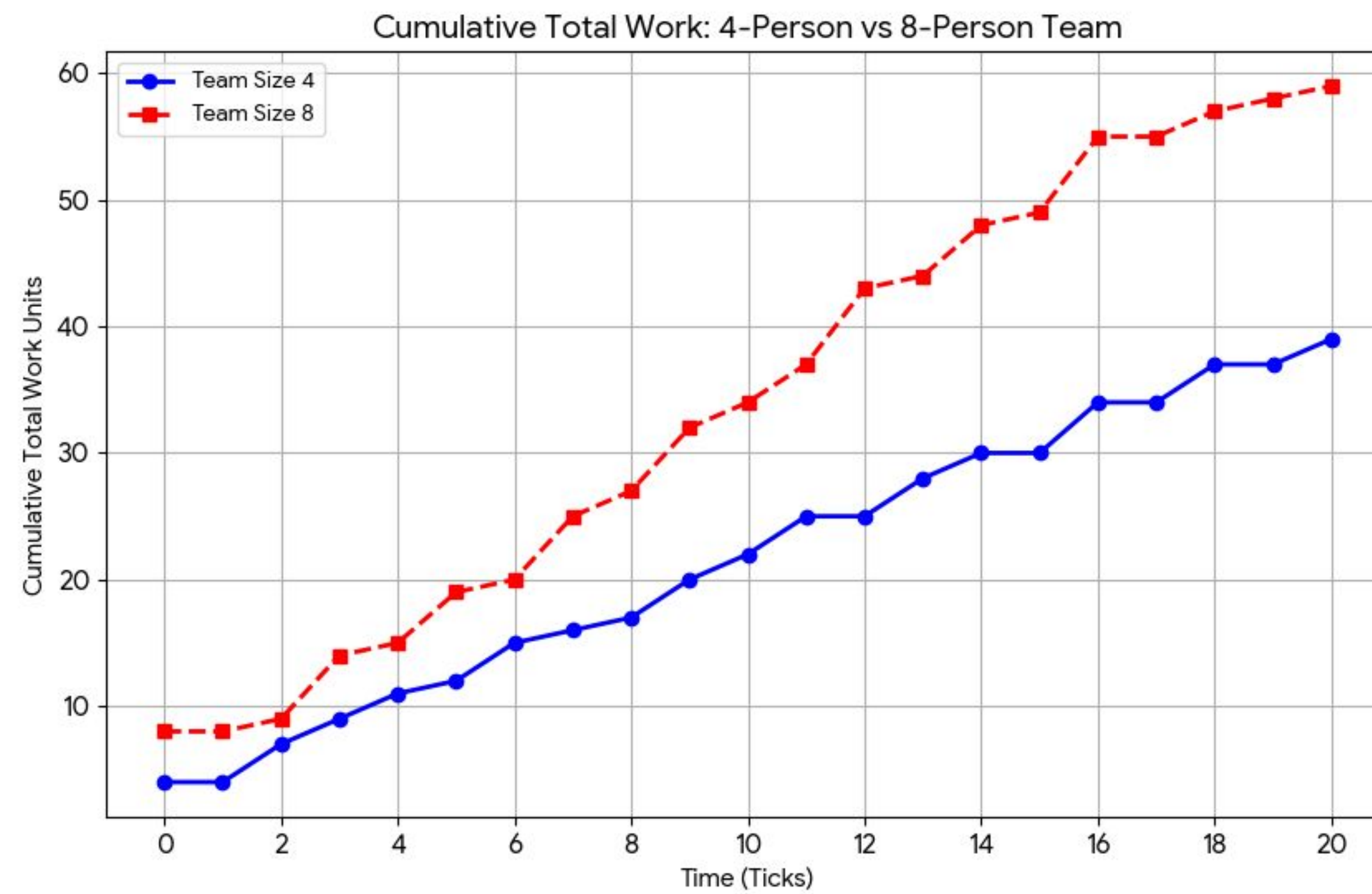
# My group work model

$$P_t = \left(1 - \frac{\text{Agents Working}_{t-1}}{N + 1}\right) \times \frac{1}{\sqrt{N/2}}$$



Time	Probability	Agents Working	Ag_1	Ag_2	Ag_3	Ag_4
0	-	4	1	1	1	1
1	0.141	0	0	0	0	0
2	0.707	2	1	0	1	0
3	0.424	1	1	0	0	0
4	0.566	2	0	0	1	1
5	0.424	1	1	0	0	0
6	0.566	4	1	1	1	1
7	0.141	0	0	0	0	0
8	0.707	3	1	1	1	0
9	0.283	0	0	0	0	0
10	0.707	4	1	1	1	1

# Is this accurate?



# Why do we care about simulations?

1. Building a simulation helps us learn about the thing we're simulating.
  - The exercise of building a simulation forces us to think really deeply about how something works.
2. Ideally, *prediction*. Can we simulate complex systems in enough detail to make predictions?
3. How is this different from modeling?







# So what next?

How do LLMs change agent-based modeling?

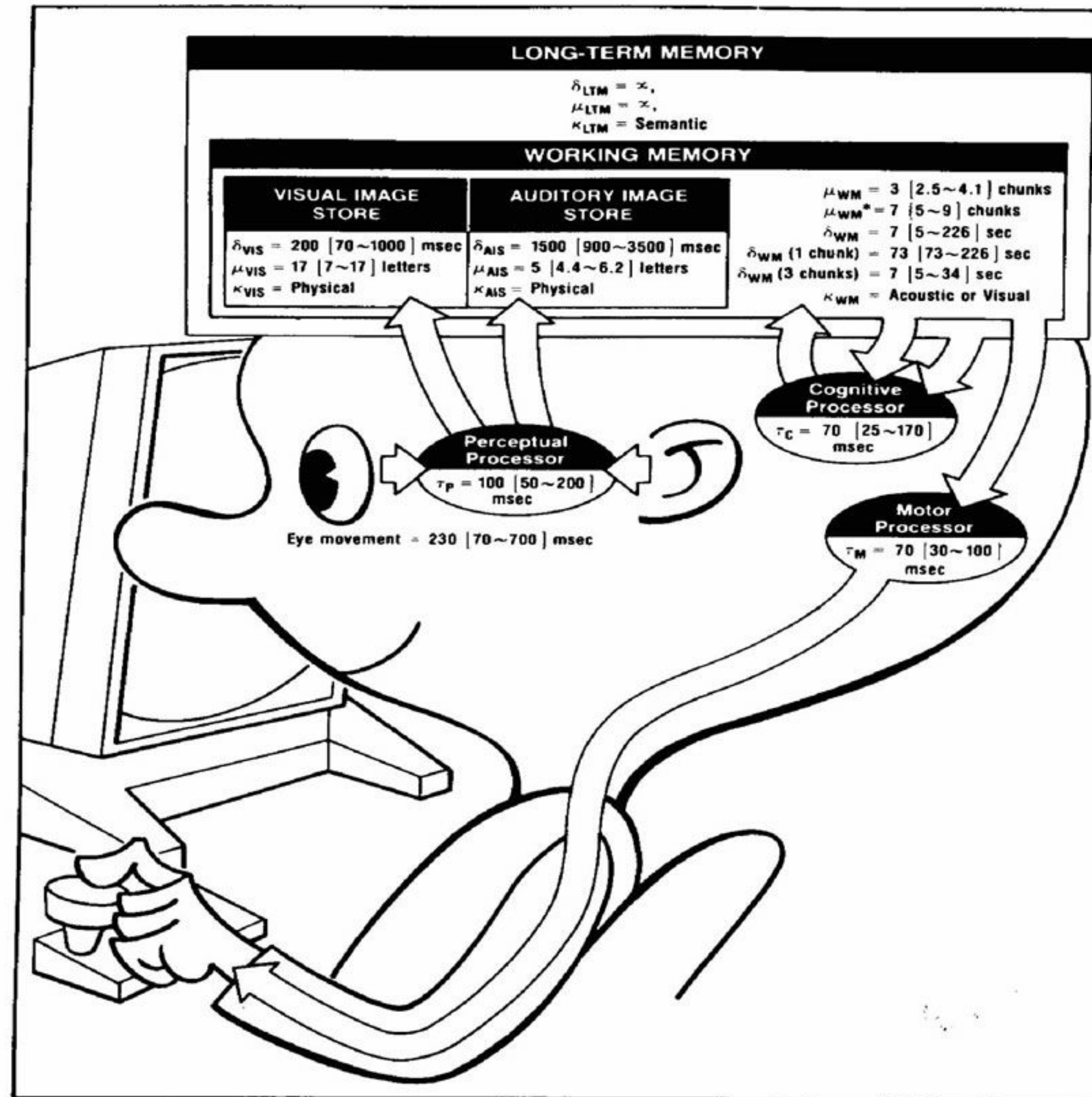
How can we simulate complex populations, now that we aren't limited to simple rules?

A simple approach would just be... Ask ChatGPT. But can we do a little better?

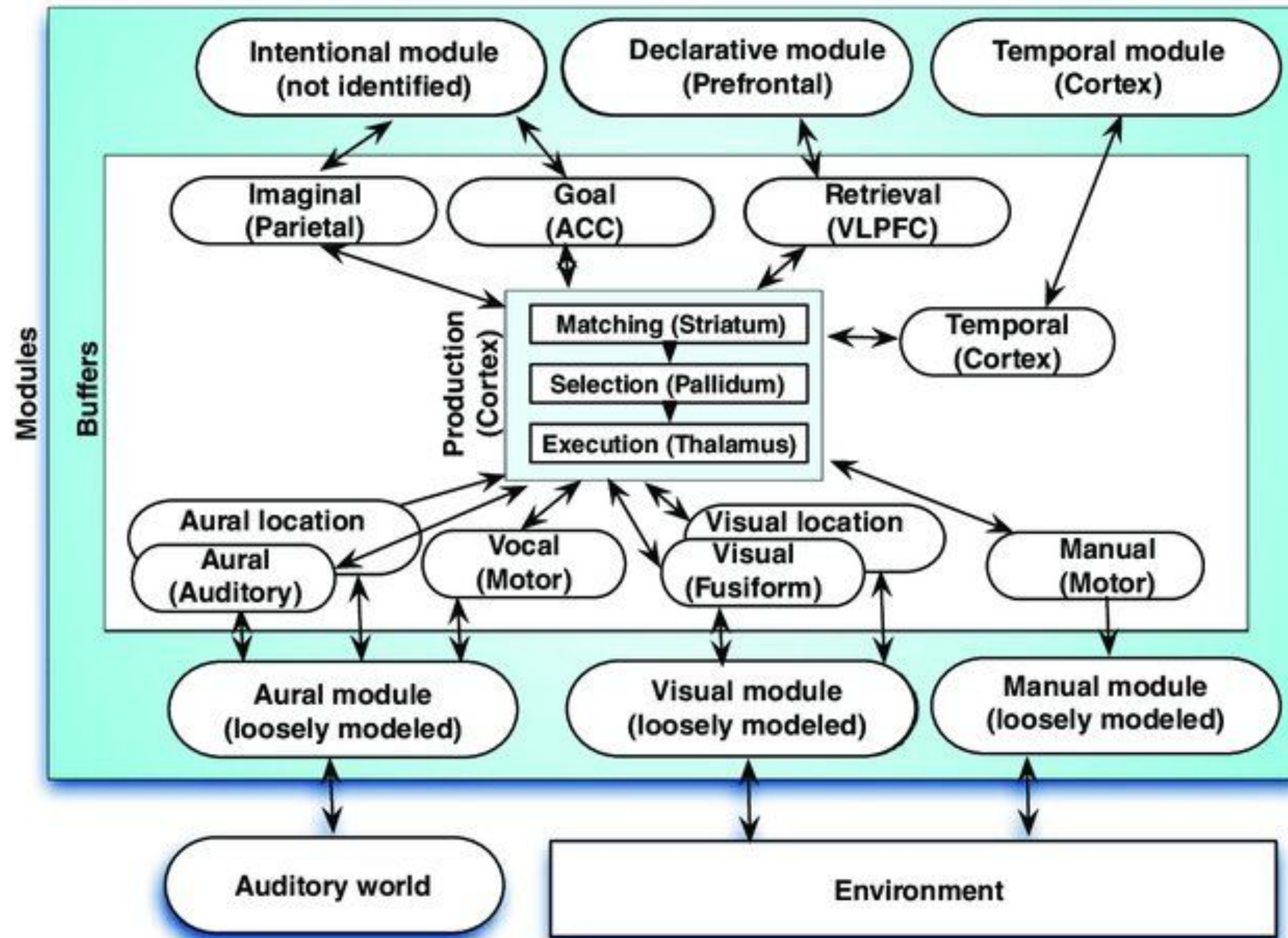
# Cognitive Architectures

We can replace simple rules with (basically) simulated brains.

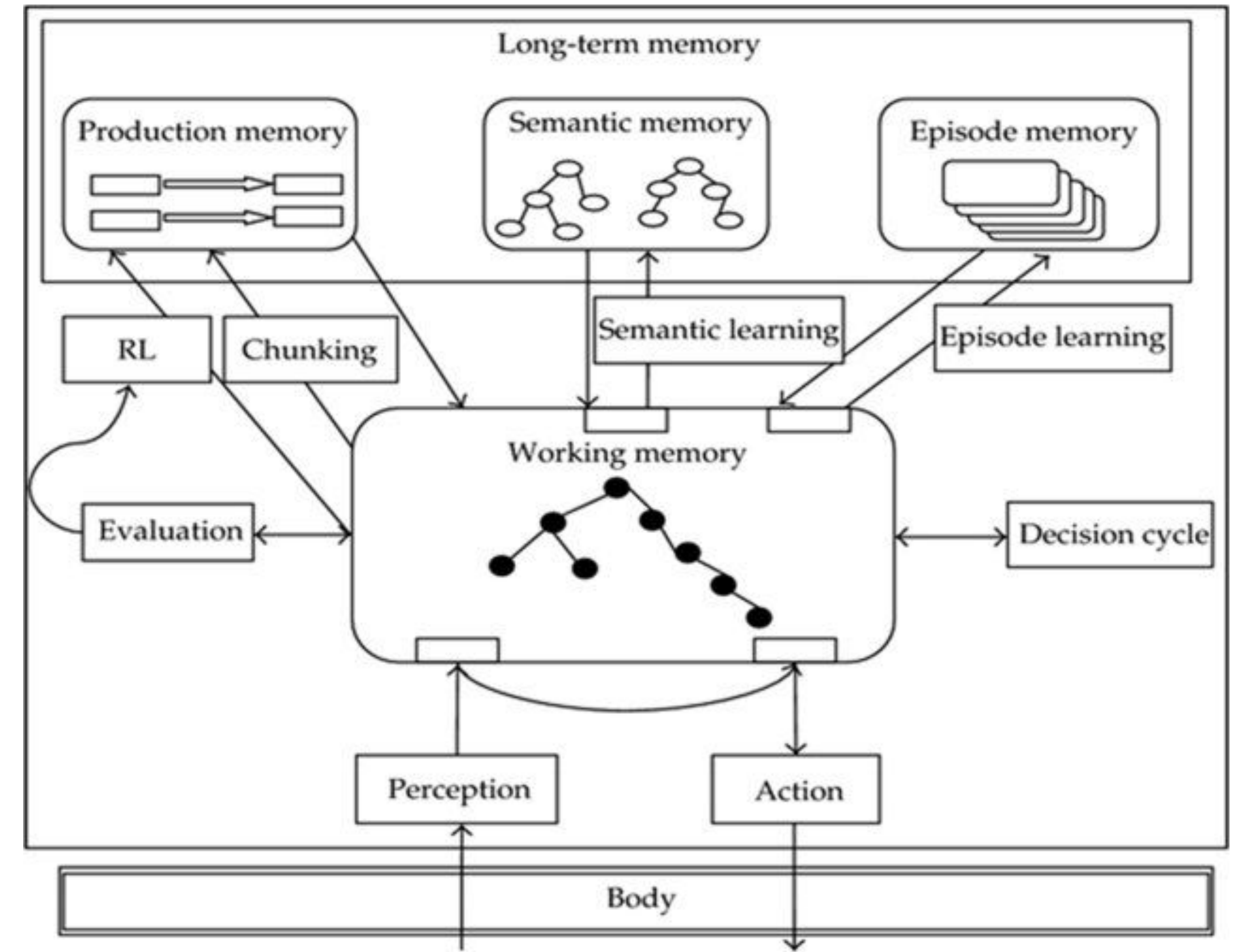
# So how do we simulate a brain?



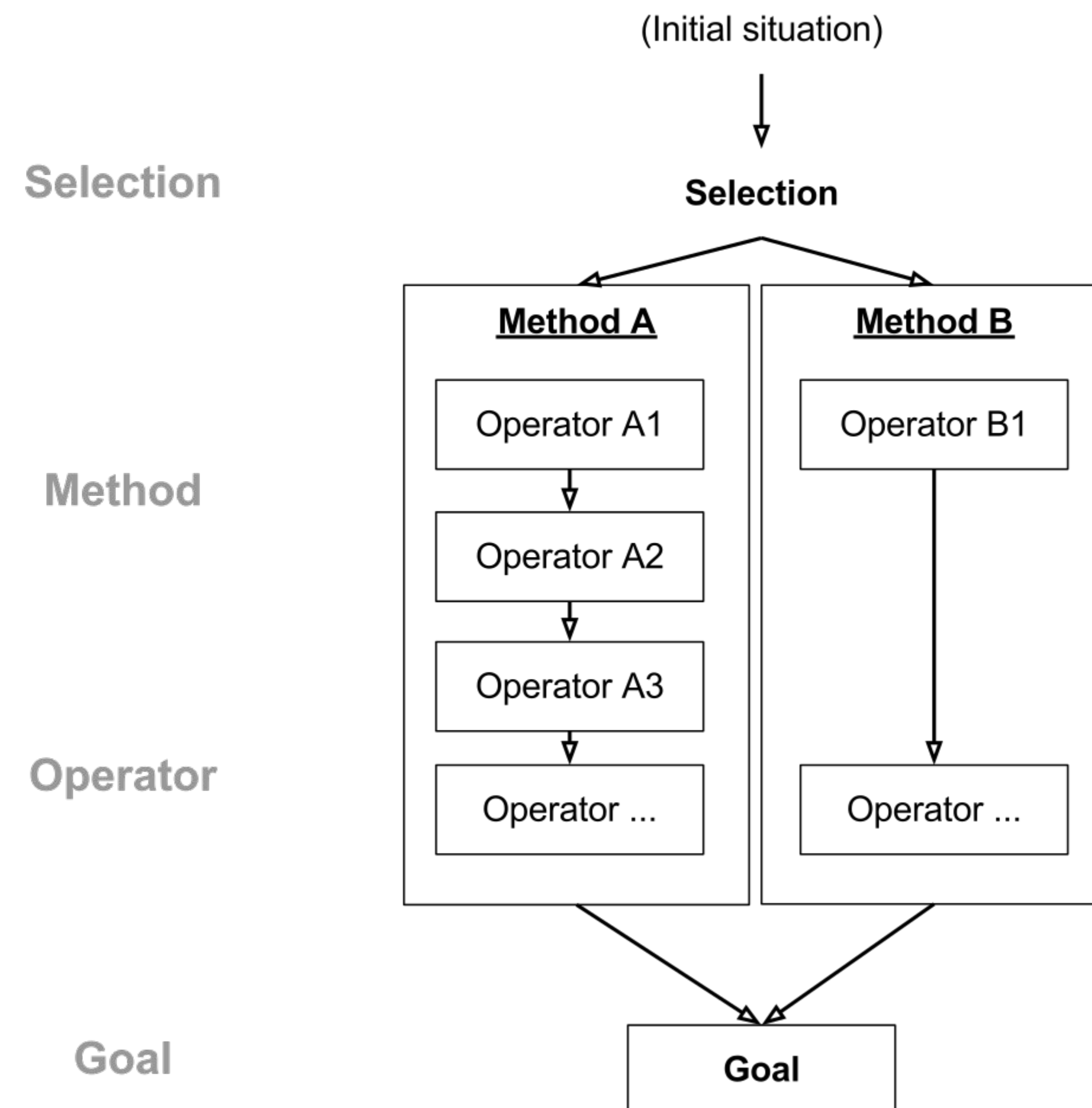




ACT-R model



SOAR model



GOMS model



# Generative Agents: Interactive Simulacra of Human Behavior

Joon Sung Park  
Stanford University  
Stanford, USA  
joonspk@stanford.edu

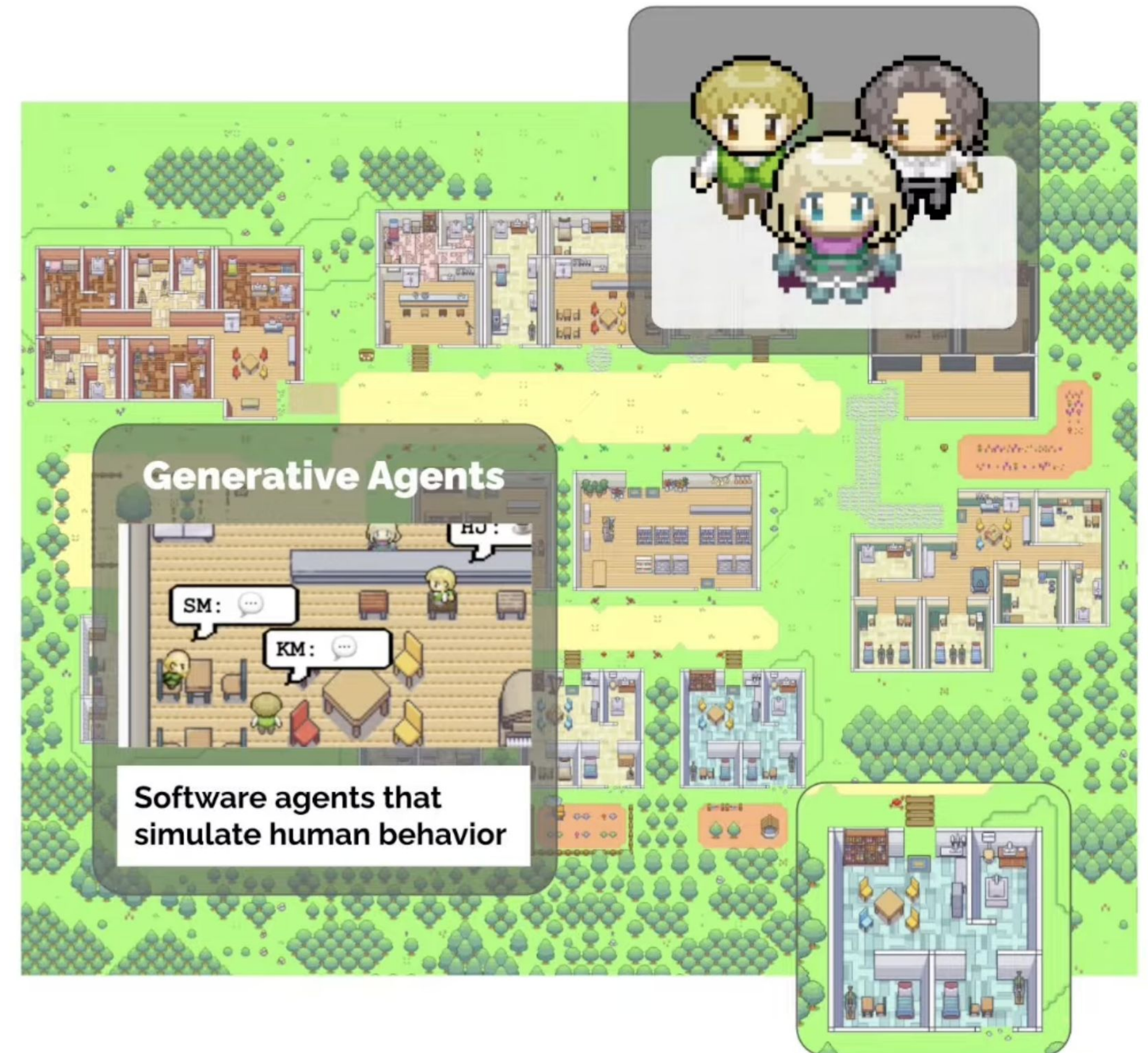
Meredith Ringel Morris  
Google DeepMind  
Seattle, WA, USA  
merrie@google.com

Joseph C. O'Brien  
Stanford University  
Stanford, USA  
jobrien3@stanford.edu

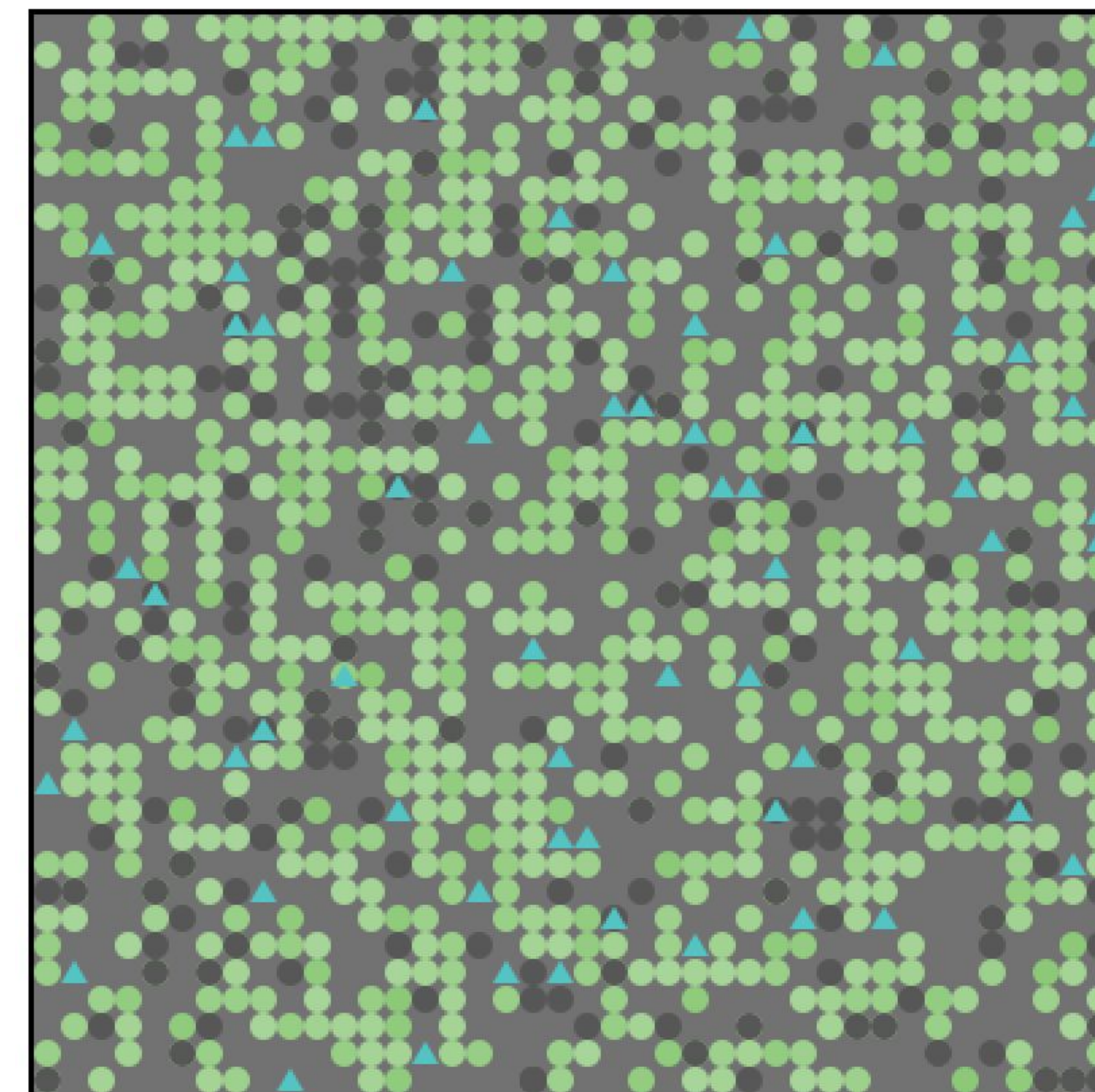
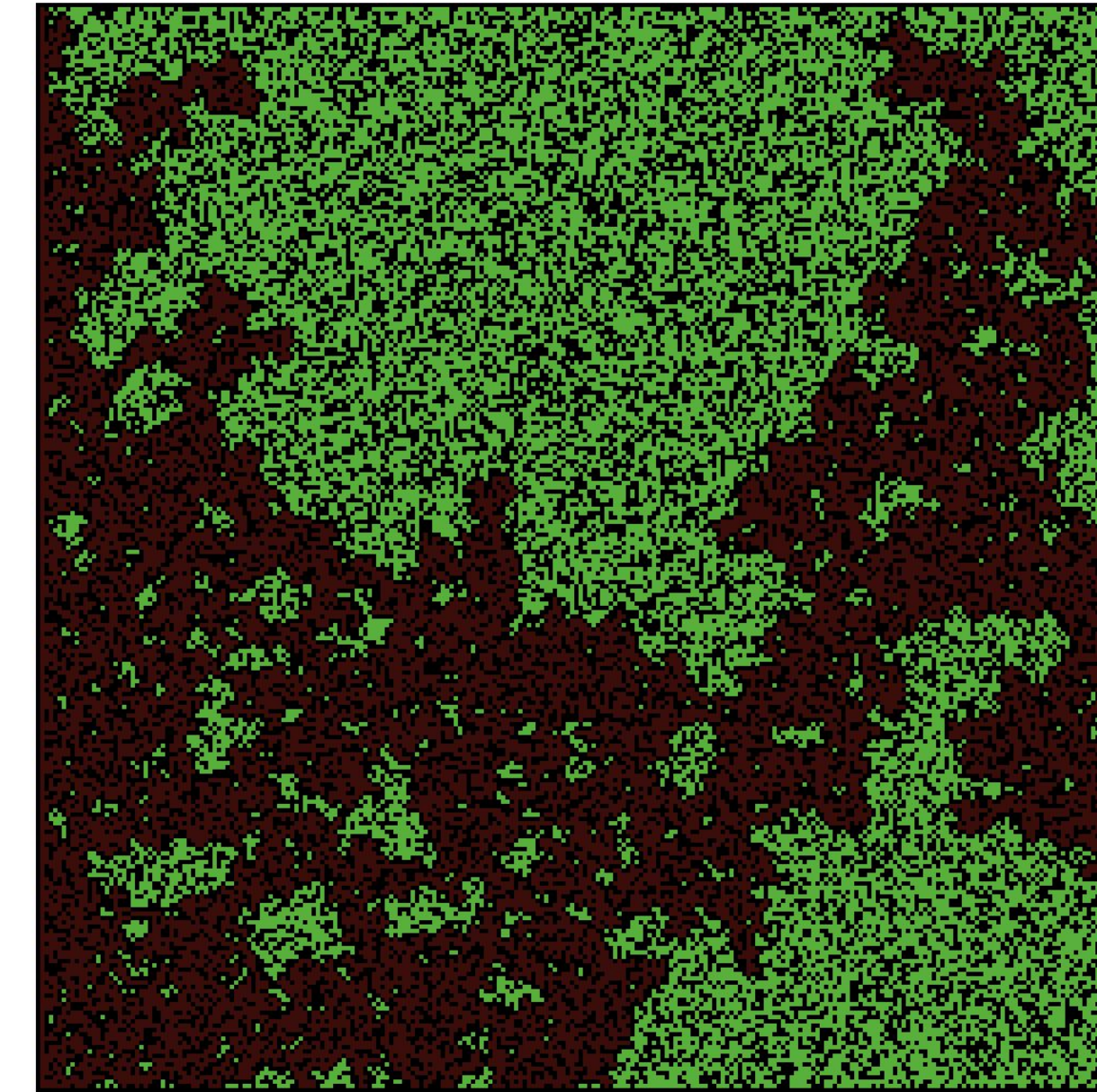
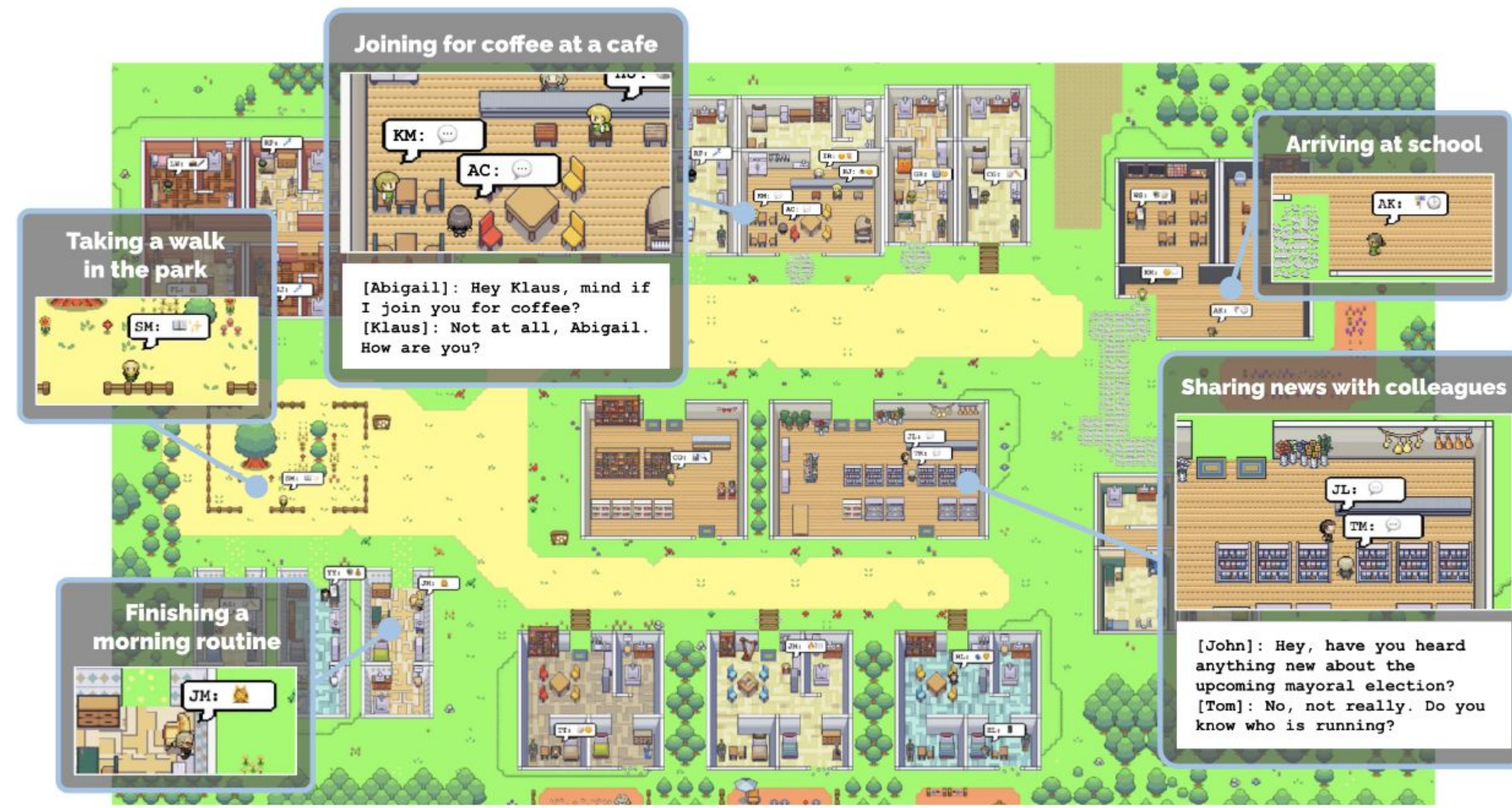
Percy Liang  
Stanford University  
Stanford, USA  
pliang@cs.stanford.edu

Carrie J. Cai  
Google Research  
Mountain View, CA, USA  
cjcai@google.com

Michael S. Bernstein  
Stanford University  
Stanford, USA  
msb@cs.stanford.edu



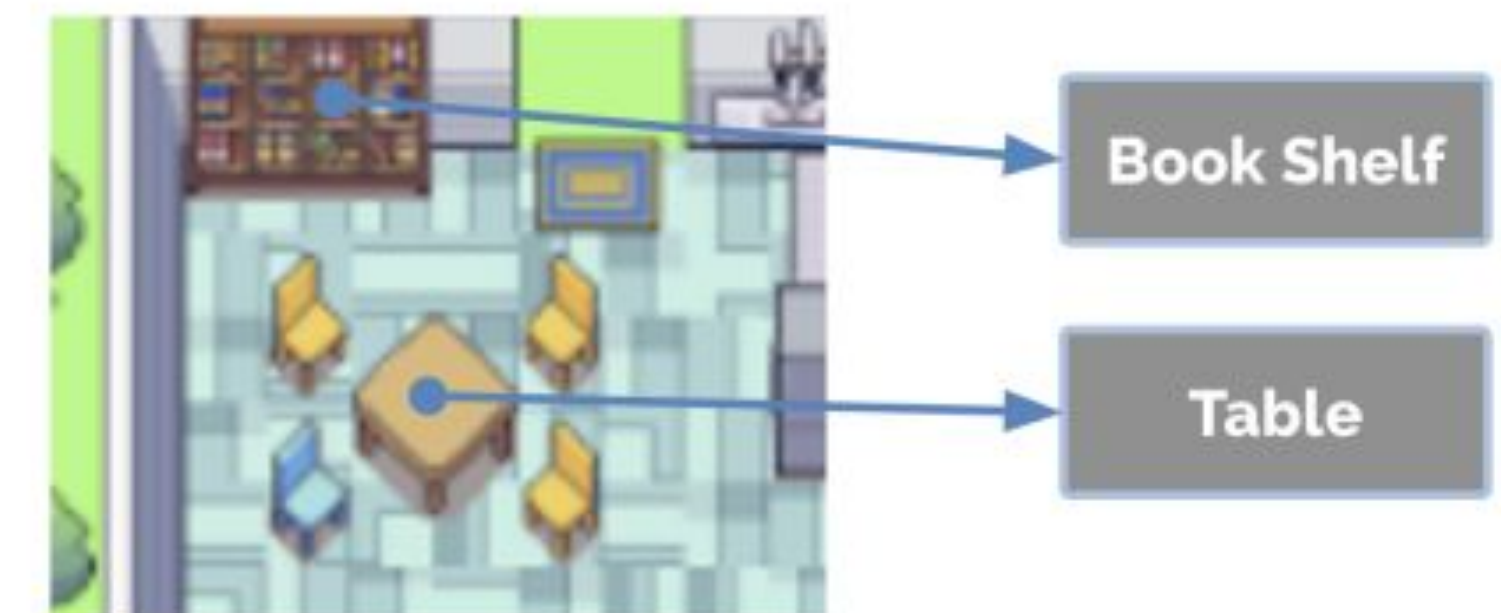






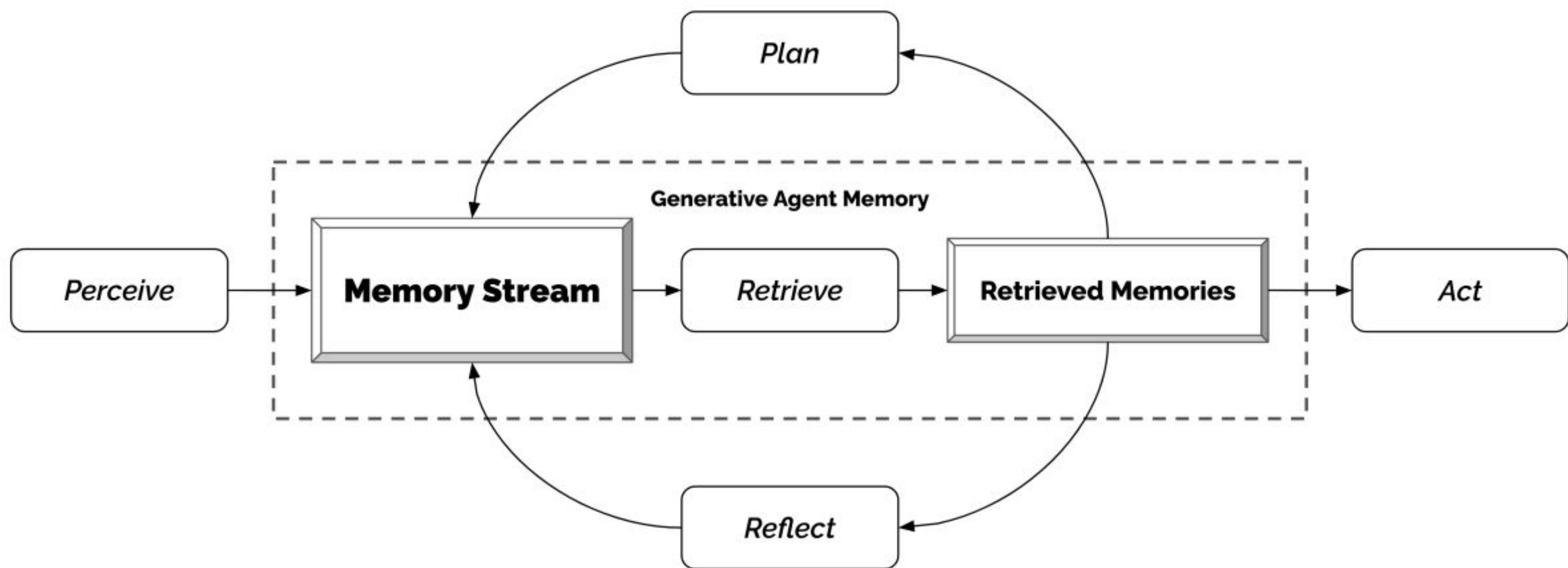


**Family House**



**Common Room**







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
------	---	------	--------	--------

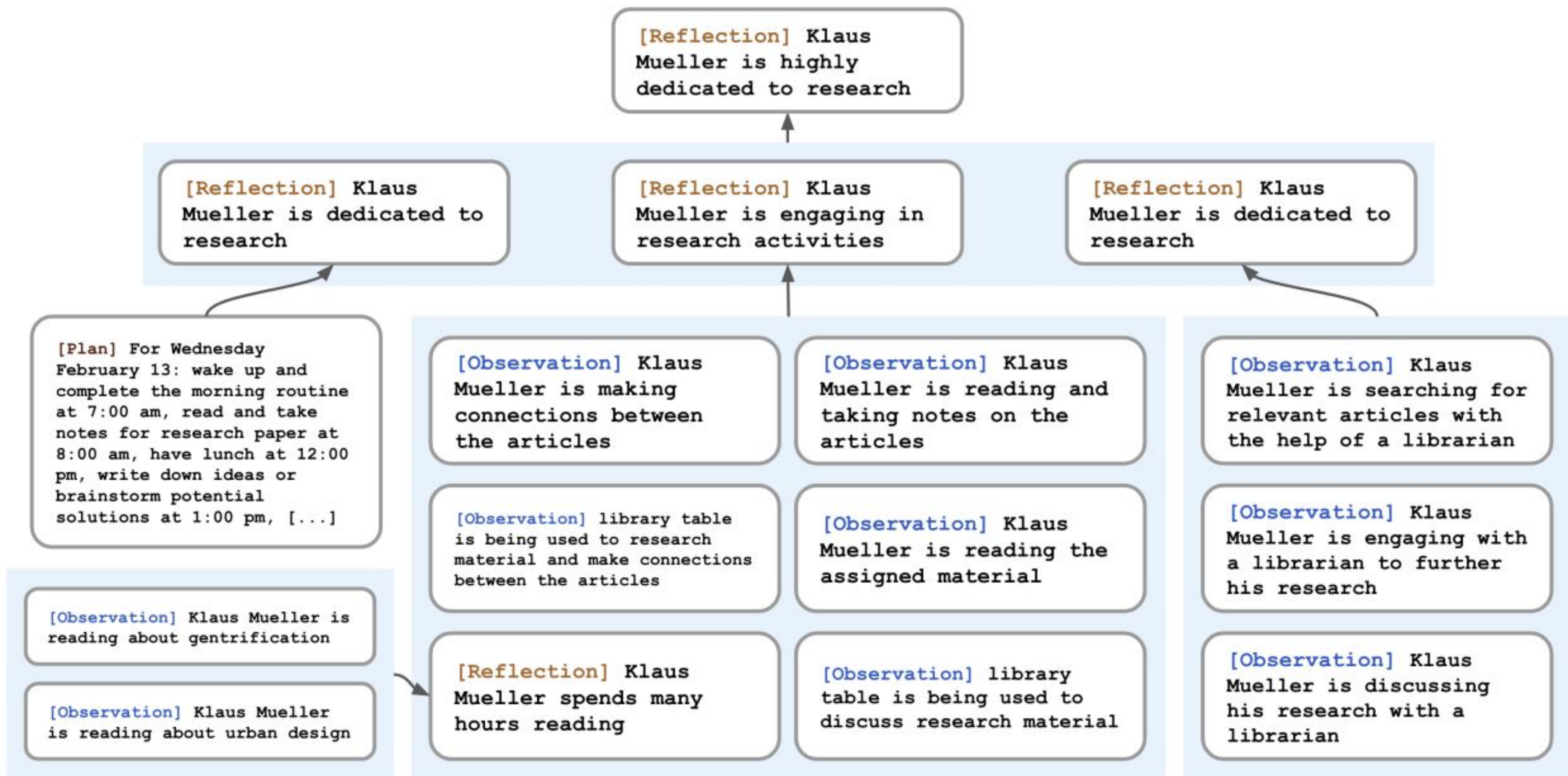
...

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



Figure 6: The memory stream comprises a large number of observations that are relevant and irrelevant to the agent’s current situation. Retrieval identifies a subset of these observations that should be passed to the language model to condition its response to the situation.





**Figure 7: A reflection tree for Klaus Mueller. The agent's observations of the world, represented in the leaf nodes, are recursively synthesized to derive Klaus's self-notion that he is highly dedicated to his research.**

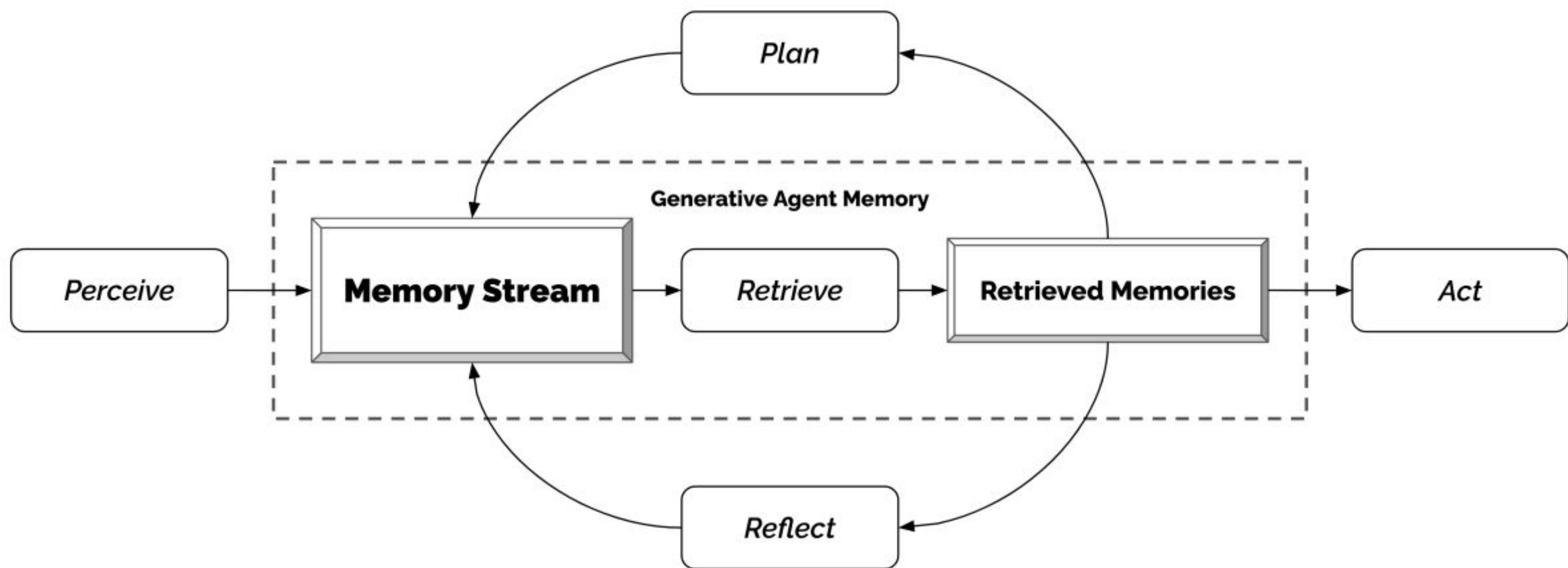


**Klaus Mueller**

**Basic information**

<b>First name</b>	Klaus
<b>Last name</b>	Mueller
<b>Age</b>	20
<b>Current time</b>	February 13, 2023, 20:58:20
<b>Current tile</b>	[114, 30]





---

# Position: LLM Social Simulations Are a Promising Research Method

---

**Jacy R. Anthis**<sup>1 2 3</sup> **Ryan Liu**<sup>4</sup> **Sean M. Richardson**<sup>1</sup> **Austin C. Kozlowski**<sup>1</sup>  
**Bernard Koch**<sup>1</sup> **Erik Brynjolfsson**<sup>2</sup> **James Evans**<sup>1 5</sup> **Michael S. Bernstein**<sup>2</sup>

Table 1: LLM social simulations must address five key challenges.

Challenge	Description	Promising Directions
Diversity	Generic and stereotypical outputs that lack human diversity	Inject humanlike variation in training, tuning, or inference (e.g., interview-based prompting, steering vectors)
Bias	Systematic inaccuracies when simulating particular human groups	Prompt with implicit demographic information; minimize accuracy-decreasing biases rather than all social biases
Sycophancy	Inaccuracies due to excessively user-pleasing outputs	Reduce the influence of instruction-tuning; instruct LLM to predict as an expert rather than roleplay a persona
Alienness	Superficially accurate results generated by non-humanlike mechanisms	Simulate latent features; iteratively conceptualize and evaluate; reassess as mechanistic interpretability advances
Generalization	Inaccuracies in out-of-distribution contexts, limiting scientific discovery	Simulate latent features; iteratively conceptualize and evaluate; reassess as generalization capabilities advance



Metaphor: “A **personal bar** where people can share their daily lives and problems with bartender and each other”

## Closed Social Space

### Channel-based Community Interaction

The screenshot shows a social media interface with a left sidebar labeled 'Communities' containing a list of channels: #BanterBar, #barstoolchatter, #DailyBarTalk, and #CheersToRealTalk. The main content area is for the #BanterBar channel, featuring a welcome message and a list of posts. A red-bordered callout box highlights the welcome message: "Welcome to #BanterBar: Your go-to spot for laughs, random thoughts, and friendly banter -- come join the fun!". A purple-bordered callout box highlights a post by GrumpyGrill: "Can we talk about how every time I binge - watch a movie series, my life feels like one giant plot twist? One minute I'm all about self-improvement, and the next I'm elbow-deep in popcorn and existential dread. Seriously, someone needs to remind me that I shouldn't relate to fictional characters this much ..".

### Smaller and Direct Chats between Users

The screenshot shows a social media interface with a left sidebar labeled 'Messages' containing a list of conversation threads. The main content area is for a direct chat with GentleGathering, featuring a list of messages. A red-bordered callout box highlights a message from TheOpenOven: "What themes are you exploring about identity in that novel?". A purple-bordered callout box highlights a message from GentleGathering: "The poetry dives into the journey of embracing one's true self."

Metaphor: “A **ballpark** with, energetic, cheerful, and joyful atmosphere, full of people, noisy but focused on the baseball game”

## Open Social Space

### Feed-based Community Interaction

The screenshot shows a social media interface with a left sidebar labeled 'Communities' containing a list of communities: BaseballBanterBrew, Home Run Hangout, Batting Bliss Zone, StadiumVibes, Joyful Base, DiamondDe, Chatter in the Stands, and Epic Ballpark Vibes. The main content area is for the Home Run Hangout community, featuring a post by CheerfulChatter20. A red-bordered callout box highlights the post: "Can't get enough of those Friday night games! Got the speakers blasting some old school jams while we cheer on the home team. Nothing beats the combo of sports and music for the ultimate hangout vibe. #GameNight". A purple-bordered callout box highlights a post by GrumpyGrill: "Can we talk about how every time I binge - watch a movie series, my life feels like one giant plot twist? One minute I'm all about self-improvement, and the next I'm elbow-deep in popcorn and existential dread. Seriously, someone needs to remind me that I shouldn't relate to fictional characters this much ..".

### Expansive Networks between Users

The screenshot shows a social media interface with a left sidebar labeled 'Home', 'Notifications', 'Messages', and 'Profile'. The main content area is for a user profile, featuring a list of followers. A red-bordered callout box highlights the list of followers: "CheerfulObserver", "FieldSideFriend", "BallparkBuzz", and "HomeRunHarmony".



Simulating Korea's future in 2050 with AI, we aim to find the 'single best option' among millions of possibilities.

Korea #1535697 is beautiful! The streets and parks are full of vitality. Various social and economic indicators look good as well.

Shall we conduct an interview with an AI resident inside the simulation? It seems the recently tested real estate policy is working well.

AI와 함께 한국의 2050년의 미래를 시뮬레이션하며, 수백만 개의 가능성 속 '단 하나의 최선'을 찾고자 합니다.

Korea #1535697

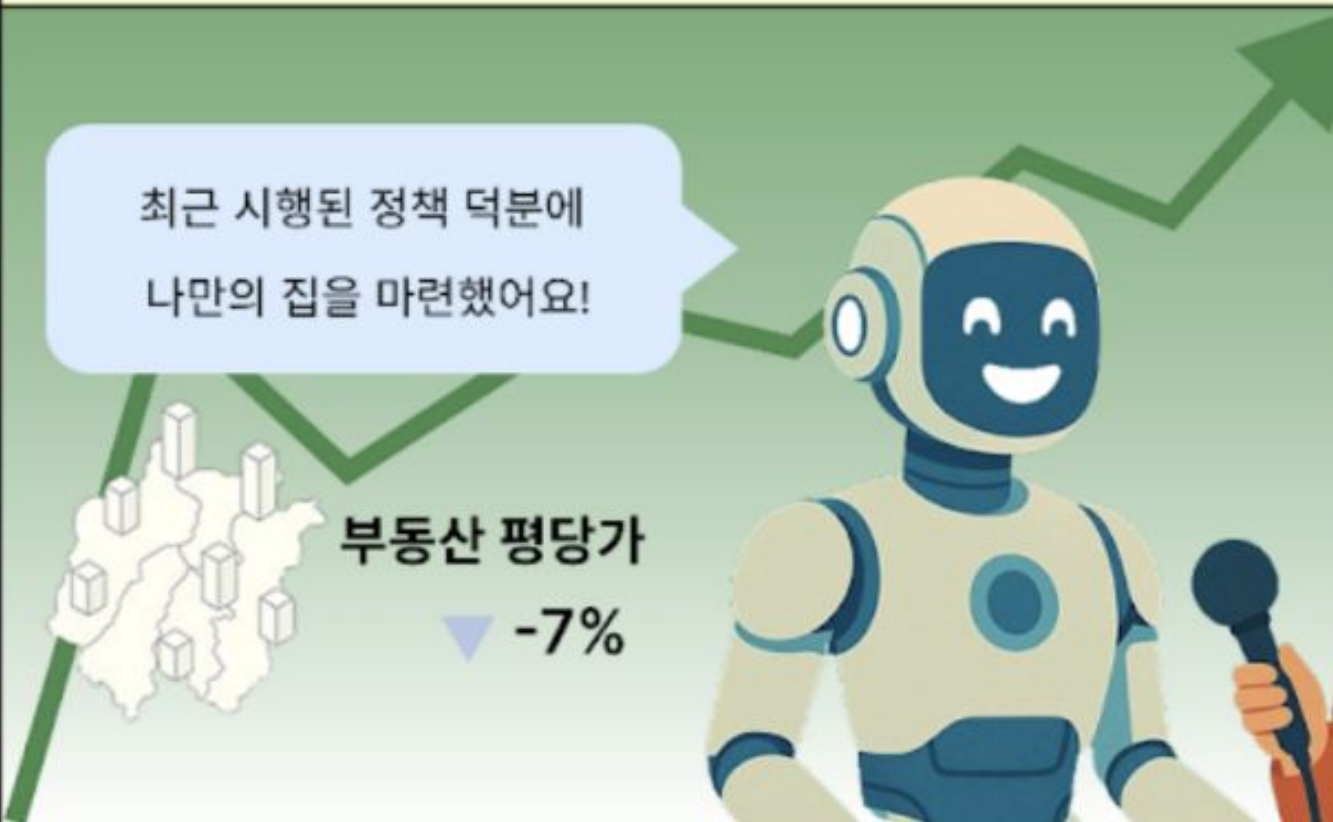


한국 #1535697의 현황을 살펴볼까요?

한국 #1535697은 아름답군요! 거리와 공원에서 활기가 넘칩니다. 사회, 경제 측면 다양한 지표들도 좋네요.




시뮬레이션 속 AI 주민과 인터뷰를 진행해볼까요? 최근 테스트한 부동산 정책이 잘 작용하고 있는 것 같습니다.



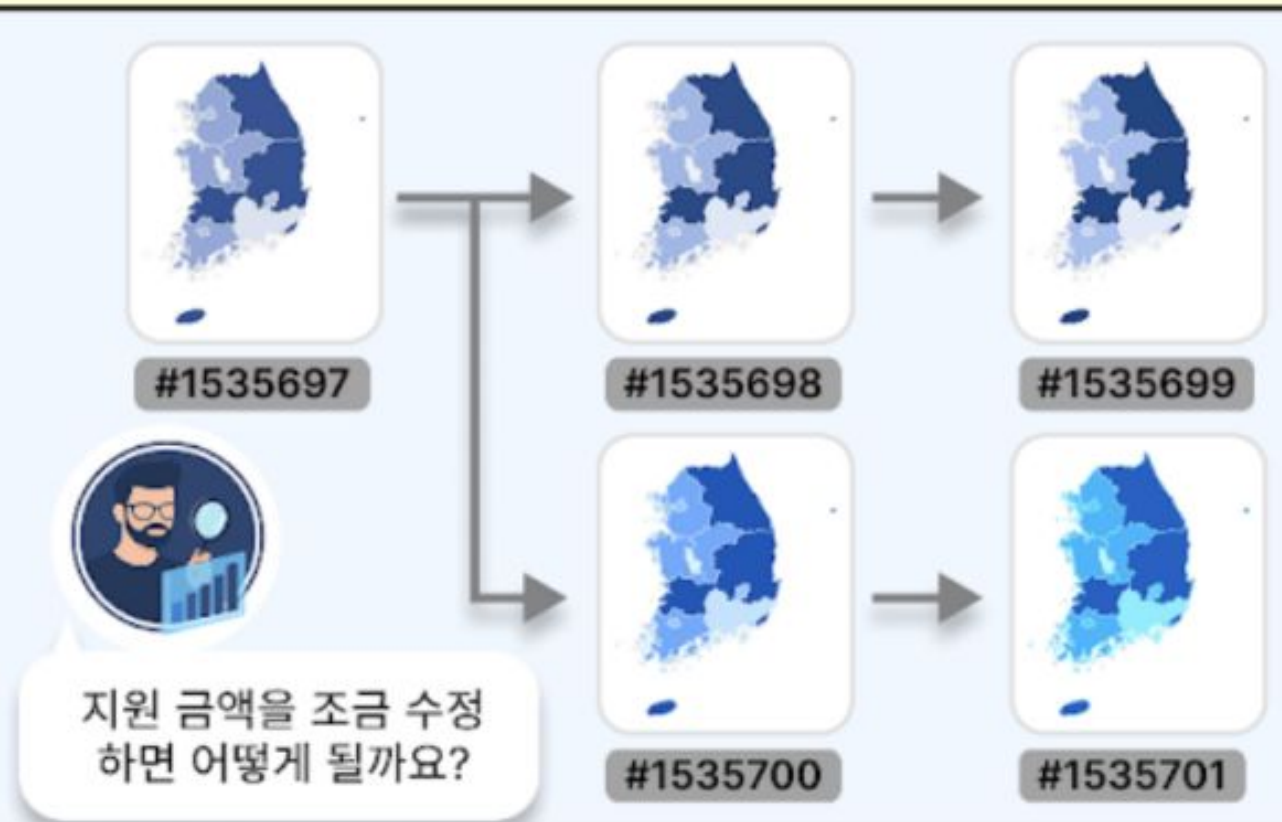
부동산 평당가  
-7%

하지만 조금 더 인터뷰 해보니, 중요한 문제가 남아있습니다. 아직 완벽한 미래는 아닌 것 같군요!



출산율  
▼ 1.0 → 0.8

시뮬레이션을 통해 다양한 정책을 테스트해보면서 최적의 결과를 찾아봅시다.



지원 금액을 조금 수정하면 어떻게 될까요?

이렇게 AI 시뮬레이션을 통해 최적의 결과를 가져오는 정책을 확인하면, 최선의 대한민국 미래를 찾을 수 있겠죠?



최고의 한국을 찾았다!

However, after interviewing a bit more, an important problem remains. It seems it isn't a perfect future yet!

Let's find the optimal result by testing various policies through the simulation

If we confirm the policies that bring the optimal results through these AI simulations, we can find the best future







# So...

To me, simulations are one of the most exciting areas  
in social computing  $\cap$  HCI

In order to be successful, we need a deep  
understanding of extremely complex human systems

**AND**

We need mastery over emerging AI technologies.

It's also a high risk field. Exciting!

$$P_t = \left(1 - \frac{\text{Agents Working}_{t-1}}{N + 1}\right) \times \frac{1}{\sqrt{N/2}}$$

Activity: Make a simple  
agent-based simulation



# Activity: Make a (less) simple agent-based simulation

<https://bit.ly/CS473Sim2>