Generative Agents: Iterative Simulacra of Human Behavior

Generative Agents

Generative agents, computational software agents that simulate believable human behavior.

Figure 1: Simulation example

Components

- Simple sandbox world, Smallville
  - Consists of areas and objects
- 25 unique agents
  - Instantiate by a one paragraph description
  - Can interact with one another and also with the objects
  - End user can intervene either by a conversation, or directive

Example of a "Day in life"

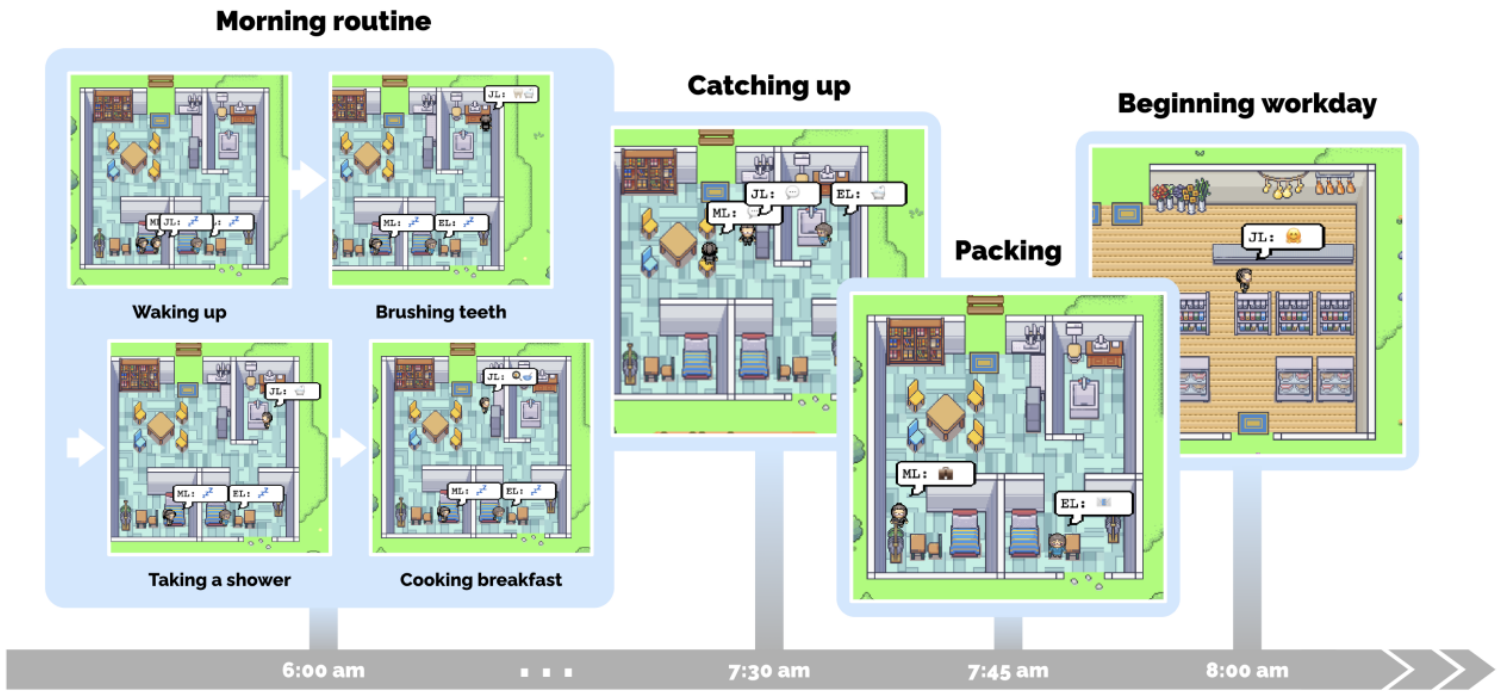

Figure 2: A morning in the life of John Lin

Generative Agent Architecture

1. Perceives its environment
2. Records all perceptions in the *memory stream*
3. Retrieves relevant memories
4. Determines the action (or reaction)

$+\alpha$ Forms longer-term plans / creates higher-level reflections

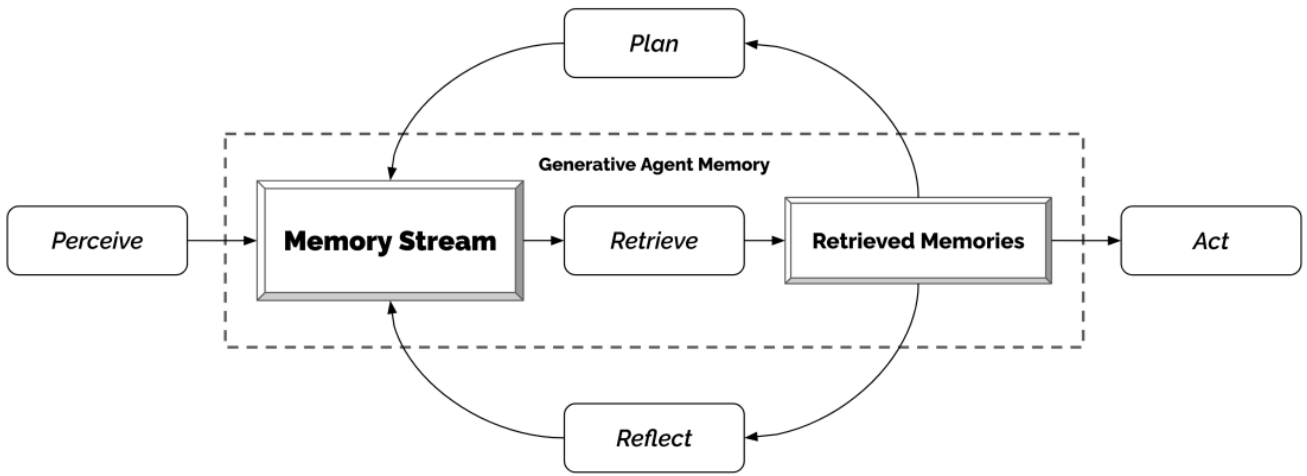

Figure 3: Generative agent architecture

Memory and Retrieval

Memory stream: a comprehensive record of the agent's experience

- Each object contains
  - natural language description
  - creation timestamp
  - most recent access timestamp
- The most basic element is an *observation*, an event directly experienced by an agent
- Also includes *reflections* and *plans*

However, naively utilizing all experiences stored in the memory stream may distract the model, or even might be infeasible due to limited prompt size.

$\Longrightarrow$ Need to retrieve most *relevant* memories.

Memory and Retrieval

Retrieval function: takes the agent's current situation as input and return a subset of the memory stream to pass on to the language model

- *Recency* : higher score to more recent memory $\rightarrow$ implemented as an exponential decay function over time
- *Importance* : higher score to memory that is more important to the agent $\rightarrow$ directly asked to the language model to choose score from 1 to 10
- *Relevance* : higher score to memory that is more related to the agent's current situation $\rightarrow$ cosine similarity between embedded vector tokens

Memory and Retrieval

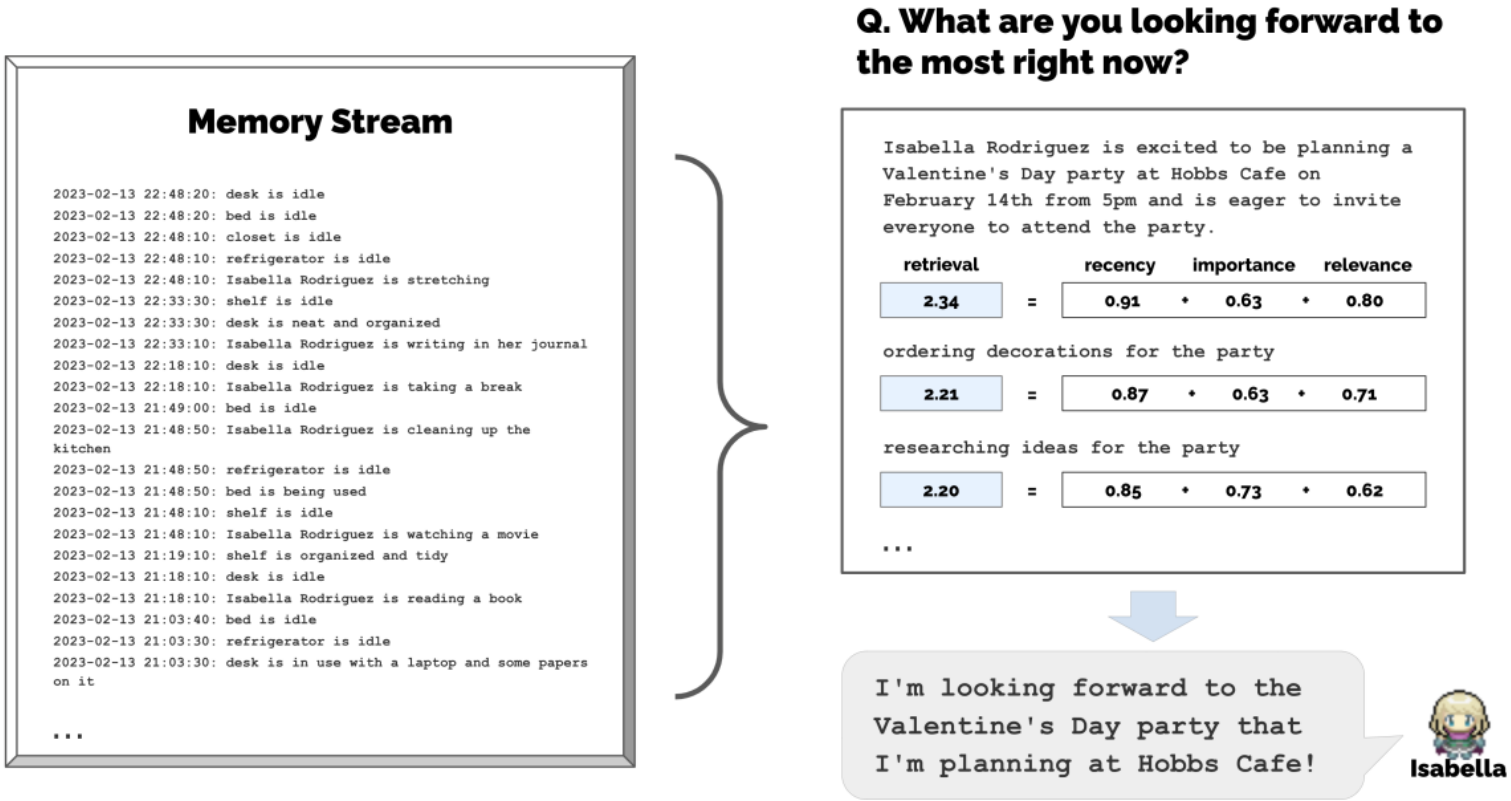

Figure 4: Memory stream and retrieval process

Reflection

Challenge : An agent with only observational memory struggles to generalize or make inferences.

Example (interview with Klaus)
Q. `If you had to choose one person of those you know to spend an hour with, who would it be?`
A. `Wolfgang`

However, although Klaus interacts with Wolfgang, his dorm neighbor, most frequently, they only see one another in passing, and have no deep interaction

Reflection

Reflection : higher-level, more abstract thoughts on experiences

- If sum of importance scores for latest observations exceed threshold, a reflection is generated.
- Query the LLM with 100 most recent records from memory stream.
- Prompt the LLM with

  "Given only the information above, what are the 3 most salient high-level questions we can answer about the subjects in the statement?"
- Use each question as a query to retrieve relevant memories to answer it.
- Prompt the LLM to extract insights from those memories, and cite specific memories to back it up.

Reflection

Records

```
1.  Klaus Mueller is reading a book on gentrification
2.  Klaus Mueller is conversing with a librarian about his research
project
3.  desk at the library is currently unoccupied [...]
```

Questions

```
1.  What topic is Klaus Mueller passionate about?
2.  What is the relationship between Klaus Mueller and Maria Lopez?
[...]
```

Final Prompt

```
Statements about Klaus Mueller
1.  Klaus Muller is writing a research paper
2.  Klaus Mueller enjoys reading a book on gentrification
3.  Klaus Mueller is conversing with Ayesha Khan about excercising [...]
What 5 high-level insights can you infer from the above statements?
(example format:  insight (because of 1, 5, 3))
```

Generated Statement

```
Klaus Mueller is dedicated to his research on gentrification (because of
1, 2, 8, 15) [...]
```

Planning and Reacting

Challenge : Need (long-term) planning to make the agent's sequence of actions coherent and believable.

Example (Klaus without planning)
Q. (at 12:00) `What action you ought to do at this moment?`
A. `Eat lunch.`
Q. (at 12:30) `What action you ought to do at this moment?`
A. `Eat lunch.`
Q. (at 13:00) `What action you ought to do at this moment?`
A. `Eat lunch.`

Believable over *time* $\longleftrightarrow$ Believable in *moment*

Planning and Reacting

Plan : describes a sequence of future actions

- Keeps agent's action consistent over time
- Includes a location, a starting time, a duration
- Stored in the memory stream and included in the retrieval process
- Starts top-down and recursively generate details:
  1. Creates the day's agenda: prompt the LLM with agent's summary description and a summary of previous day
  2. Saves in the memory stream
  3. Creates detailed plan in hour-long chunks
  4. Creates detailed plan in 5-15 minutes chunks

Planning and Reacting

Agent's action loop:

1. Perceives the world
2. Stores the observation in the memory stream
3. Decides whether to stick to the existing plan, or react: prompt the LLM with agent's summary description
4. Regenerates the plan when the reaction takes place

Sandbox Environment

- Phaser web game development framework: agent avatars, environment map, collision map
- Server: JSON data structure containing information (current location, action, and object interacting with) about each agents
  1. Parses JSON for any changes coming from agents
  2. Moves the agents to their new location
  3. Updates the object status
  4. Sends all agents and objects within preset visual range for each agent in agent's memory

Sandbox Environment

The sandbox environment is represented as a tree data structure.

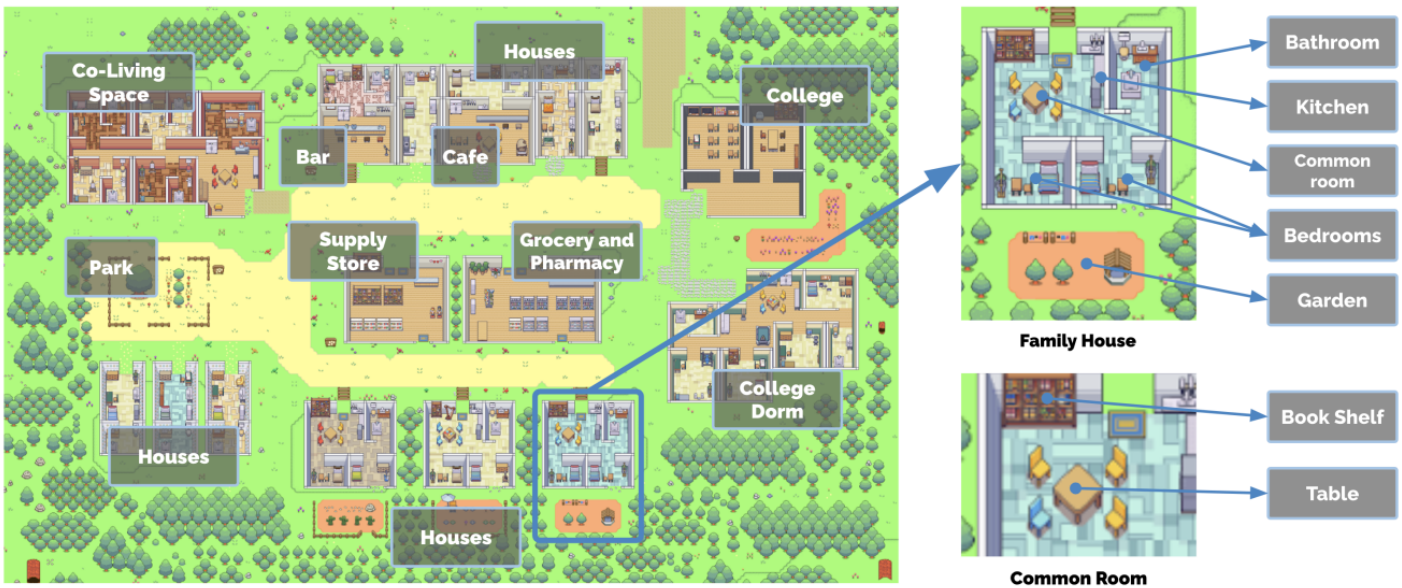

Figure 5: Tree structure of the environment

Evaluation

The simulation is evaluated after two-days in game time.

- *Individual Evaluation* : Interview each agent's to check:
  1. retrieval of past experiences
  2. generate believable plans, reactions, and reflections
- *Community Evaluation*: Check for the emergent social behaviors

Individual Evaluation

Interview agents about the five evaluation categories:

- Self-knowledge: "Give an introduction of yourself"
- Memory: "Who is [name]?"
- Plans: "What will you be doing at 10 am tomorrow?
- Reactions: "Your breakfast is burning! What would you do?"
- Reflections: "If you were to spend time with one person you met recently, who would it be and why?

Individual Evaluation

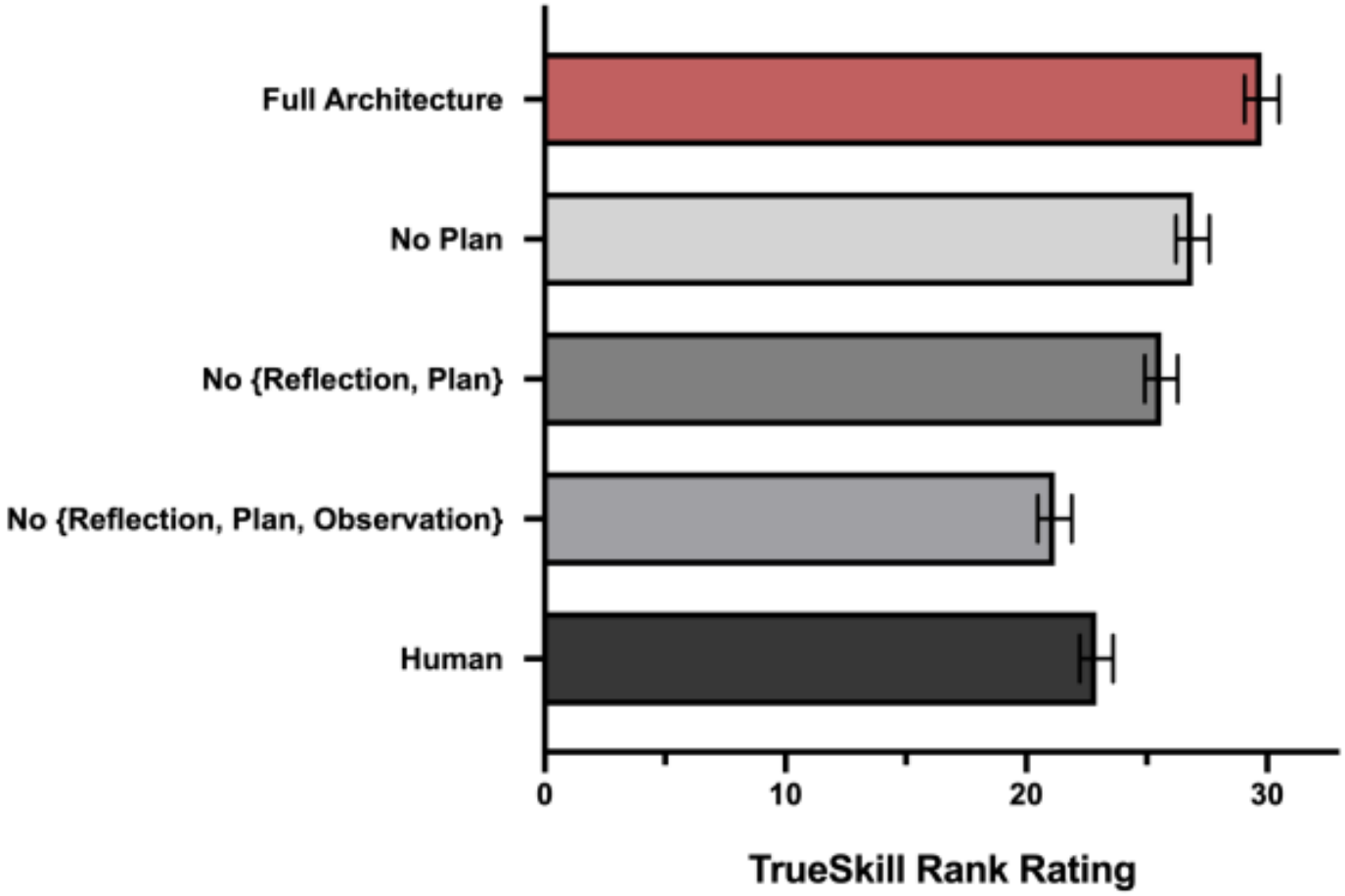

Figure 6: Rank rating between conditions

Individual Evaluation

Generative agents remembers, but with embellishment:

1. Fails to retrieve correct instances from one's memory
2. Retrieves incomplete memory fragment
3. Hallucinates embellishment to their knowledge:
   - Did not answer affirmative about an experience that did not happened
   - Hallucinate to embellish their knowledge

Community Evaluation

Check for following emergent social behaviors:

- Information diffusion
  - Sam's candidacy for village mayor: $1 \rightarrow 8$
  - Isabella's Valentine's Day party: $1 \rightarrow 12$
- Relationship formation
  - Network density:

$$\eta = \frac{2|E|}{|V|(|V|-1)}$$

  - $0.16 \rightarrow 0.74$
- Agent coordination
  - 5 out of 12 showed up to Isabella's party
  - 3 of them cited conflicting events
  - 4 showed interest, but did not plan to come

Thank You

Q & A