

Generative Agents

Interactive Simulacra of Human Behavior

Park, O'Brien, Cai, Morris, Liang, Bernstein | Stanford University & Google Research | UIST 2023

Believable agents that live, remember, and coordinate



Figure 4: Five agents gather at Isabella's Valentine's Day party — planned and coordinated entirely by the agents themselves.

- 25 generative agents populate a small sandbox town (Smallville) for two full simulated days
- From a single seed idea — “Isabella wants to throw a party” — agents spread invitations, form new relationships, and show up together at the right place and time
- Human evaluators rated the full architecture far more believable than every ablation and a human-crowdworker baseline (effect size $d = 8.16$)

Why believable agents are hard to build

- For four decades, researchers have wanted computational agents that act as believable proxies of human behavior (game NPCs, social simulation, prototyping tools)
- Rule-based methods (finite-state machines, behavior trees) require heavy manual authoring
- Reinforcement learning succeeds mainly in narrow, reward-defined games
- Large language models can produce plausible behavior at a single moment, but struggle to stay coherent over a long, changing, social world

Architecture: memory, reflection, planning

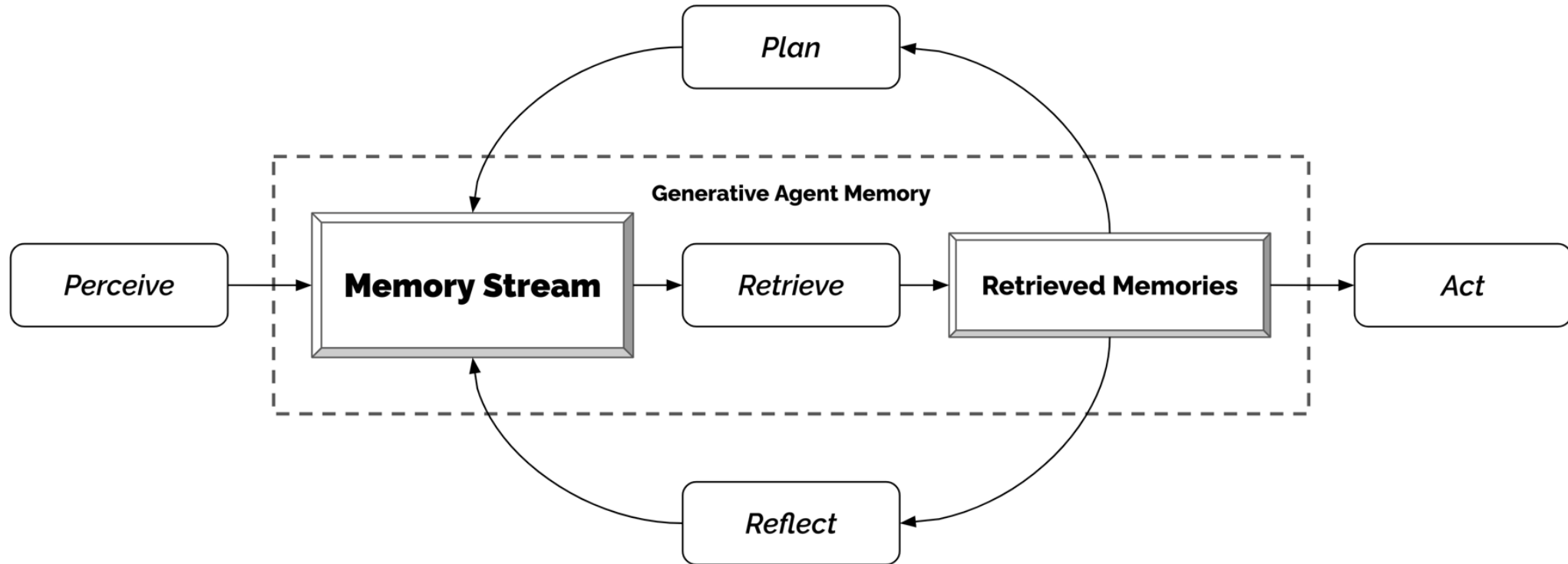


Figure 5: Agents perceive their environment; retrieved memories inform planning and reaction; plans and reflections feed back into the memory stream.

Smallville: a sandbox town of 25 agents

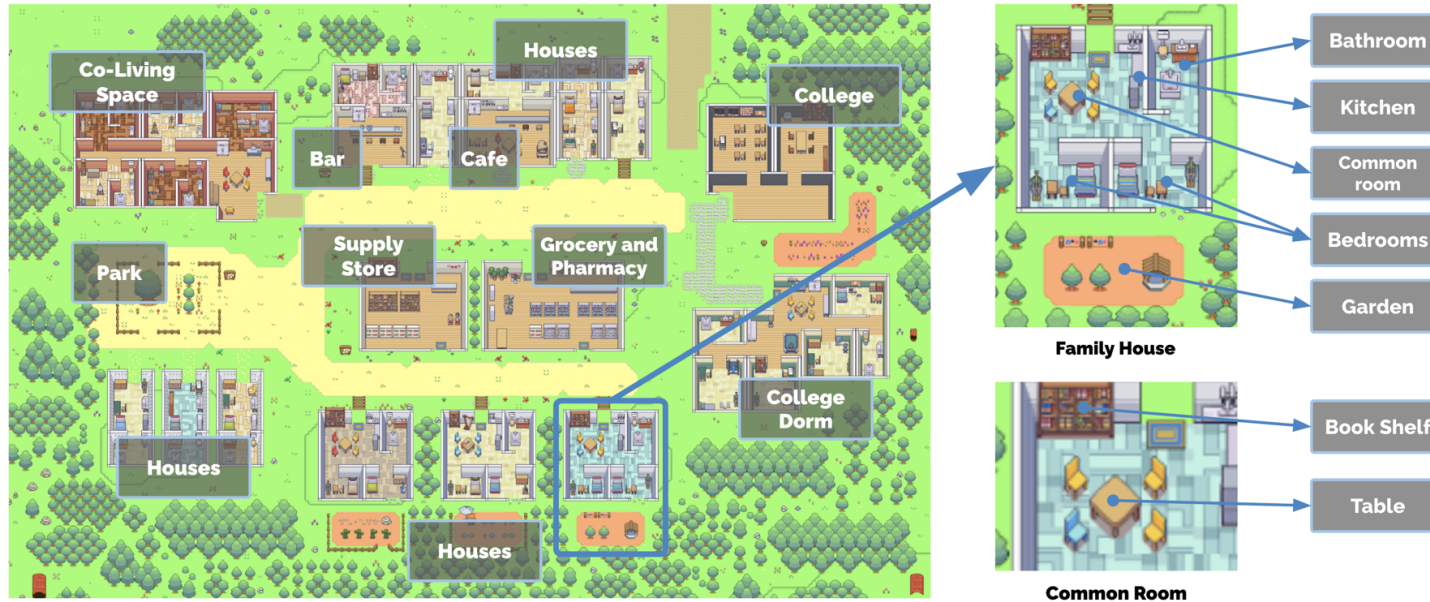


Figure 2: The Smallville sandbox world, with areas labeled.

- A Sims-like world with houses, a cafe, bar, park, school, and stores
- The world is represented as a tree: whole world → areas → objects, rendered into natural language
- Each agent maintains its own partial view, updated only as it explores

Scoring memories: recency, importance, relevance

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

2.34

=

recency

0.91

+

importance

0.63

+

relevance

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

...

Recency: exponential decay since last access

Importance: the LLM rates poignancy 1–10 when a memory is created

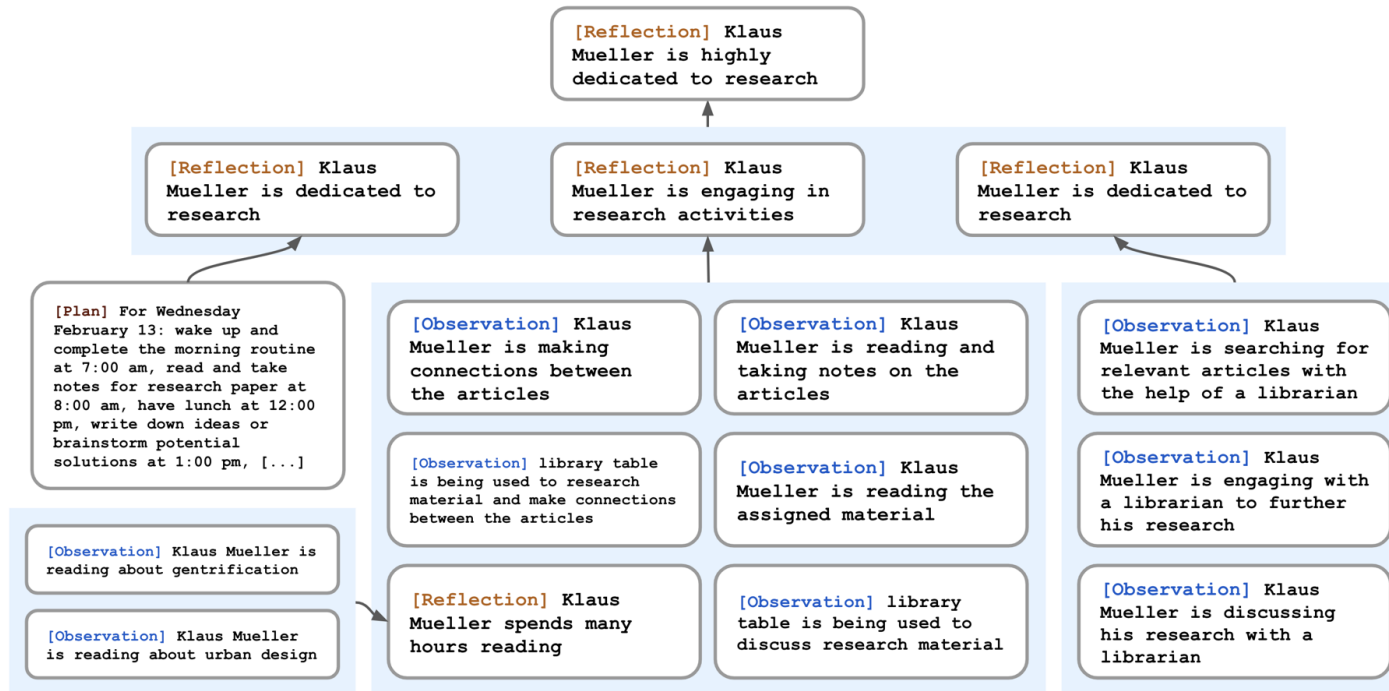
Relevance: cosine similarity between memory and query embeddings

The three scores are combined to select which memories enter the prompt

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

Isabella

From raw observations to abstract plans



- Reflection: periodically synthesizes recent memories into higher-level insights (e.g. “Klaus is dedicated to research”), with citations back to evidence
- Planning: a day is first sketched in broad strokes, then recursively refined into hour-level and 5–15 minute actions
- Agents can react to new observations and replan mid-stream

Figure 7: A reflection tree — observations synthesized into Klaus's higher-level self-knowledge.

Interviews show the full architecture is most believable

- Agents were interviewed on self-knowledge, memory, plans, reactions, and reflection
- 100 participants ranked 5 conditions by believability: full architecture, 3 ablations, and a human crowdworker
- Full architecture scored highest (TrueSkill $\mu = 29.89$); removing memory, planning and reflection scored lowest ($\mu = 21.21$)
- Effect size vs. the memory-less baseline: $d = 8.16$ (all pairwise differences significant, $p < 0.001$)

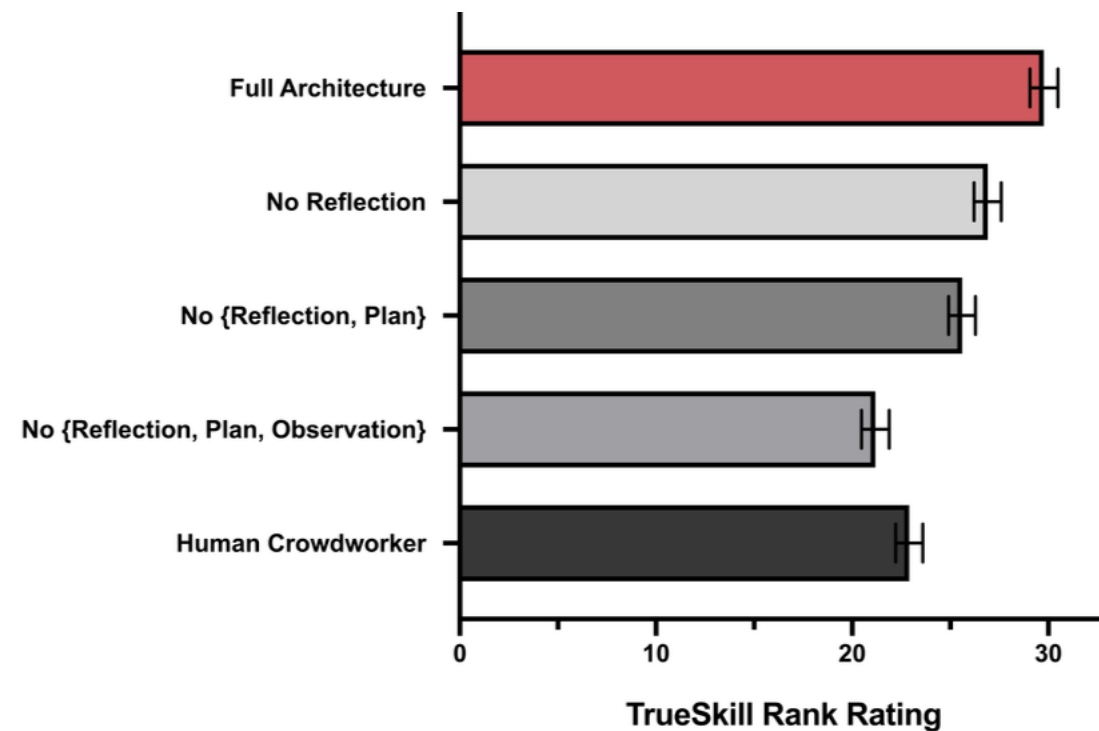
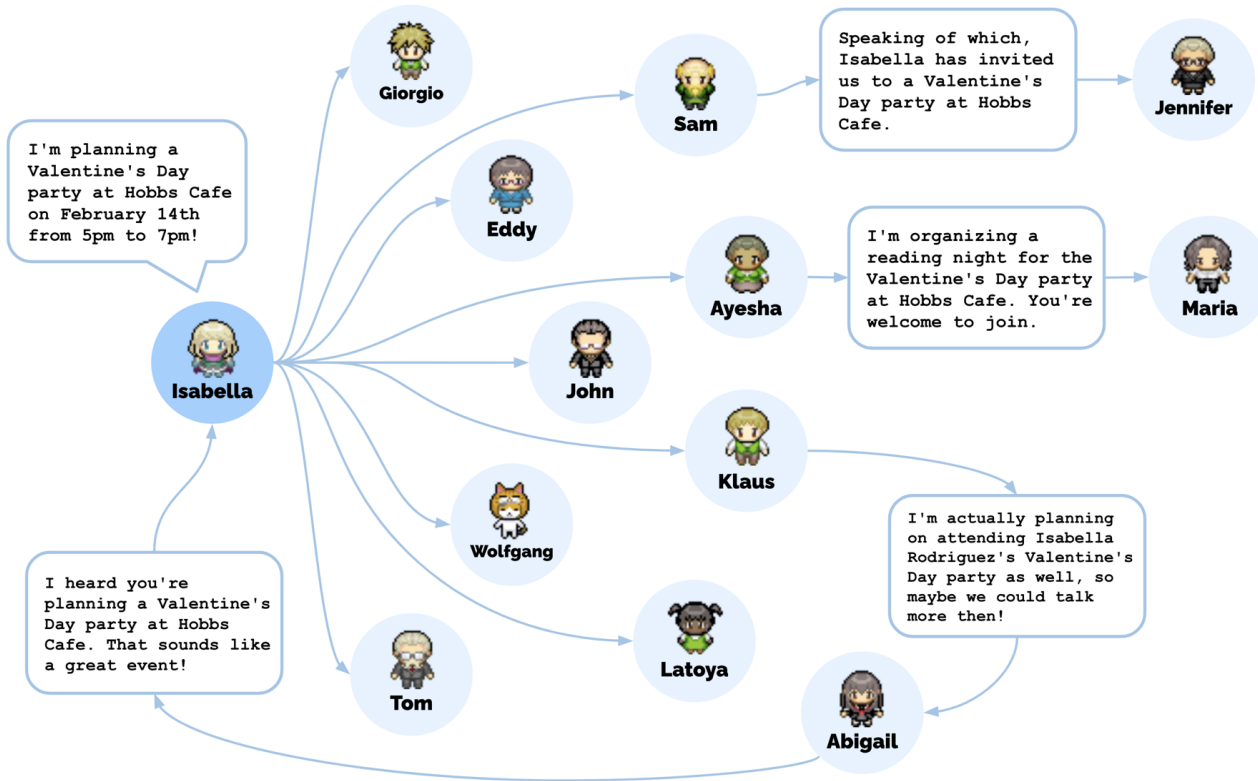


Figure 8: The full generative agent architecture produces more believable behavior than the ablated architectures and the human crowdworkers. Each additional ablation reduces the performance of the architecture.

Emergent diffusion, relationships, and coordination



- Awareness of Sam's mayoral candidacy: 4% → 32% of agents
- Awareness of the Valentine's Day party: 4% → 52% of agents
- Relationship network density: 0.167 → 0.74
- 5 of 12 invited agents actually attended the party
- Only 1.3% of relationship-awareness answers were hallucinated

Figure 9: Diffusion path of Isabella's party invitation across 12 agents.

Limitations and risks

- Memory retrieval is imperfect — agents sometimes miss or partially recall relevant memories
- Occasional hallucinated embellishments drawn from the LLM's own world knowledge
- Spatial norms not captured in language can be misjudged (e.g. entering a closed store)
- Instruction-tuned dialogue can feel overly formal and agreeable
- Ethical concerns: parasocial attachment, error propagation, and misuse for deepfakes or persuasion — the authors recommend disclosure, value alignment, and audit logging

CONCLUSION

Memory + reflection + planning yield believable agents

- An LLM alone is not enough — the surrounding architecture is what produces long-term coherent, believable behavior
- Emergent social dynamics (diffusion, relationships, coordination) arose from simple seeds, without any scripting
- Opens a path toward social prototyping, role-play, games, and human-centered design tools