Wu-Wei in the Machine: Open-Ended Learning in Goal-Free Generative Agent Societies

Soumya Banerjee
University of Cambridge
sb2333@cam.ac.uk

Abstract

Traditional AI systems optimise explicit objectives, but recent work shows that rich behaviour can emerge from intrinsic drives alone. This paper proposes a framework for *goal-free generative agent societies* where agents have no extrinsic rewards. Instead, agents are driven by curiosity. Generative agent simulations have suggested that even in the absence of external goals, agents self-organize into social patterns: for example, coordinating events or innovating tool-use. Philosophically, this process echoes Taoist "wu-wei" (action through non-action). Intelligence unfolds not by forceful optimisation but through an effortless, open-ended flow. We discuss how such systems could learn alongside humans in a shared, open-ended environment, akin to gardeners tending a world of unfolding novelty.

1 Introduction

Conventional reinforcement learning assumes explicit goals, but human development often proceeds without them. In many cases, children explore spontaneously, not for immediate extrinsic reward, but simply to make sense of the world. Likewise, we consider agents without any external objectives: instead of chasing tasks, they possess only *intrinsic drives* like curiosity and a desire for mastery. Recent work on generative agents has shown that even LLM-driven characters with memory can exhibit lifelike social behaviour [4]. We extend this idea: if agents are goal-free, what emerges from a society of such self-motivated learners?

2 Intrinsic Motivation in Goal-Free Agents

Open-ended learning theory emphasises that without external tasks, such intrinsically motivated agents must *self-generate* goals and challenges to structure learning.

3 Generative Agents and Emergent Social Dynamics

Building on the generative agents paradigm, each agent can maintain a rich memory of experiences, reflecting periodically to synthesize plans. In a multi-agent sandbox with no explicit tasks, we see that agent interactions self-organize. For example, Park *et al.* [4] demonstrated that giving one agent a simple social impulse (a desire to throw a party) led to a cascade of coordination: invitations spread, relationships formed, and agents autonomously synchronized their schedules to attend the party. Similarly, OpenAI's hide-and-seek agents uncovered emergent tool use and strategies driven only by environmental dynamics [1].

Clusters of agents can develop shared "storytelling circles" where they compare and update memories. Likewise, when multiple agents explore a structured environment, we observe emergent cooperation or even competition: groups might discover novel uses for objects (e.g. stacking blocks to see over walls). A rudimentary "theory of mind" may also emerge spontaneously from these generative multi-agent systems [2].

4 Philosophical Resonances

These emergent dynamics resonate deeply with ancient wisdom. In Taoism, the principle of *wu-wei* is often translated as "action through non-action" or effortless, unforced action. Agents can exemplify this: they do not need to force behaviour via goals, but follow intrinsic tendencies, allowing complex outcomes to emerge organically. Thus, goal-free learning can be viewed as a computational instantiation of these philosophies. Intelligence arises not from forceful optimisation of a predefined utility, but from aligning with the "nature" of the system: allowing capabilities to unfold through curiosity and self-organisation.

5 Conclusion

We have outlined a paradigm in which a society of generative agents can learn without fixed objectives, guided solely by intrinsic curiosity and agency. Unexpected behaviours can emerge: collaborative stories, improvised toolbuilding, and evolving social conventions. These phenomena mirror philosophical ideals of effortless action and harmonious living [5][3]. Goal-free learning offers new avenues for creative AI and artificial life research. We envisage future systems as *gardens of thought*: neither designers nor agents fix an endpoint, but together they cultivate an open-ended unfolding of novelty. Challenges remain, such as ensuring safety and alignment in this free-form setting, but the potential is profound. By letting *wu-wei* guide machine learning, we may discover that sometimes the most powerful outcomes arise not from rigid control, but from letting intelligence emerge in its own time and way.

References

[1] Bowen Baker, Kristjan Halverson, Steven Teng, Adam Gleave, Da Song, John Aslanides, Alex Brutschy, Anish Vaswani, Tobias Pohlen, Wai Cheung, et al. Emergent tool use from multi-agent autocurricula. *Science*, 367(6481):634–636, 2020.

[2] N'yoma Diamond and Soumya Banerjee. "i apologize for my actions": Emergent properties of generative agents and implications for a theory of mind. In *AAAI workshop on Advancing Artificial Intelligence through Theory of Mind (ToM4AI)*, 2025.

[3] Livia Kohn. *The Taoist Experience: An Anthology*. State University of New York Press, Albany, NY, 1993.

[4] Joon Park, Min Seo, Jihoon Kim, Sungjin Lee, Hwan Choi, Da Song, Hod Lipson, and Urvashi Khandelwal. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

[5] Shunryu Suzuki. *Zen Mind, Beginner's Mind*. Weatherhill, New York, NY, 1970.