CraftUtopia: A LLM-based Multi-Agent System for Collaborative Construction in Minecraft

Demonstration Track

Wanli Fu*
Shanghai Artificial Intelligence Laboratory
Northwestern Polytechnical University
Shanghai, China

Hao Li*
Shanghai Artificial Intelligence Laboratory
Northwestern Polytechnical University
Shanghai, China

Siyue Ren*
Shanghai Artificial Intelligence Laboratory
Northwestern Polytechnical University
Shanghai, China

Chenxi Xing*
Yunnan University
Kunming, China

Yang Chen
Shanghai Artificial Intelligence Laboratory
Shanghai, China

Chen Chu
Yunnan University of Finance and Economics
Kunming, China

Zhen Wang†
Northwestern Polytechnical University
Xi'an, China

Shuyue Hu†
Shanghai Artificial Intelligence Laboratory
Shanghai, China

ABSTRACT

LLM-based multi-agent construction is a growing research area but prior works still suffer from low success rates for construction, reliance on predefined templates, and poor scalability as the number of agents increases. To address these challenges, we present *CraftUtopia*, an LLM-based multi-agent system (MAS) that constructs 3D structures in *Minecraft* from a single 2D reference image. CraftUtopia operates in two stages: **Design**, which converts the image into a Minecraft-compatible 3D blueprint, and **Build**, which decomposes the blueprint into spatially disjoint subtasks for parallel execution. CraftUtopia scales efficiently via two mechanisms: (i) **hierarchical coordination**, which organizes agents in a manager–foreman–worker hierarchy to separate responsibilities, and (ii) **skill acquisition**, which distills recurring action sequences into a shared skill library to reduce repeated LLM replanning. Across three representative builds, CraftUtopia achieves 100% success over five trials using only 2D inputs, scales effectively with more workers, and exhibits emergent human-like behaviors. A full demo is available at: https://github.com/craftutopia-demo/CraftUtopia.

KEYWORDS

Multi-agent Construction; Cooperation; Large Language Model

ACM Reference Format:
Wanli Fu, Hao Li, Siyue Ren, Chenxi Xing, Yang Chen, Chen Chu, Zhen Wang, and Shuyue Hu. 2026. CraftUtopia: A LLM-based Multi-Agent System for Collaborative Construction in Minecraft: Demonstration Track. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/PUIJ7087

*These authors contributed equally to this work.
†Corresponding authors: hushuyue@pjlab.org.cn, w-zhen@nwpu.edu.cn

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/PUIJ7087

1 INTRODUCTION

Multi-agent systems for construction have attracted significant interest [5, 11, 12], driven by the demand for safer, more cost-effective, and increasingly automated urbanization. Meanwhile, recent advances in large language models (LLMs) have shown strong potential for coordinating multi-agent systems [3, 7, 8, 15, 18], motivating increasing attention to LLM-based collaborative construction [4, 6, 17].

In this paper, we introduce *CraftUtopia*, an LLM-based multi-agent system for collaborative construction in Minecraft. Given a *single* 2D reference image of a target structure (e.g., a pyramid), CraftUtopia designs, plans, decomposes, and coordinates a team of agents to reproduce the corresponding 3D construction in the game. This setting surfaces several key challenges that remain underexplored in prior LLM-based multi-agent work. First, whereas most existing systems coordinate LLM agents in relatively closed domains (e.g., math problem solving or coding), we study an open-world environment with long-horizon tasks, partial observability, and non-stationary dynamics [1, 9, 14]. Second, unlike template- or blueprint-dependent approaches [10, 16], CraftUtopia builds directly from a single reference image without assuming access to predefined construction templates. Third, as prior approaches show, increasing the number of agents does not necessarily improve performance or efficiency: larger teams invite increasingly frequent social friction, turning coordination into the primary bottleneck rather than a source of efficiency [2, 4, 13, 17].

CraftUtopia defines four distinct agent roles: an architectural designer, a project manager, foremen, and workers, with the number of foremen and workers being variable. The system operates in two stages. In the **Design** stage, the designer transforms a 2D image into a 3D blueprint compatible with Minecraft. In the **Build** stage, the project manager decomposes the blueprint into spatially disjoint subtasks for foremen; each foreman plans for their respective workers, who then carry out the construction tasks. To maximize efficiency as the number of agents grows, we introduce a **hierarchical agent system** ("manager → foreman → worker") alongside a **skill acquisition process**. The skill acquisition process distills successful action sequences of workers into a shared skill library

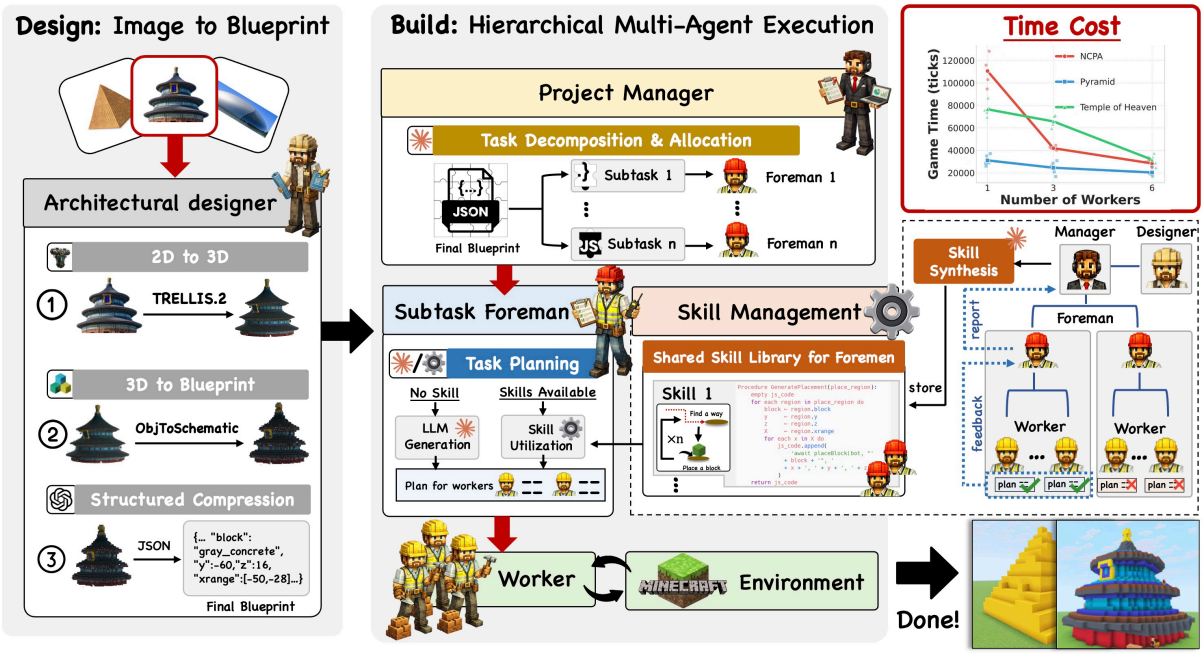

Figure 1: CraftUtopia: an LLM-based MAS that enables agents to construct arbitrary 3D building from a single 2D images.

that foremen can leverage. By invoking these learned skills instead of always relying on LLM-based re-planning foremen can accelerate construction and reduce errors. Furthermore, this process fosters cross-team knowledge sharing: foremen report new skills to the project manager, who, in turn, disseminates the knowledge across the system.

Results. We implement CraftUtopia using Claude-Sonnet-4.5 as the base model, along with additional tools for the Designer, and evaluate it on three representative builds: the Egyptian Pyramids, as well as the Temple of Heaven and National Centre for the Performing Arts (NCPA) in China (shown in the top left corner of Fig. 1). We compare CraftUtopia with the recent MINDcraft [17] study on LLM-based construction in Minecraft. In 5 trials, MINDcraft fails to complete the Temple of Heaven and NCPA even when provided with a complete 3D blueprint (compared to our single 2D image input). For the simpler Pyramids build, it succeeds in only 2 out of 5 trials. In contrast, CraftUtopia achieves 100% success across all three builds in five trials using only a single 2D image. Even when MINDcraft succeeds, its build time is at least 2× longer than CraftUtopia's *worst-case* runtime for the same build. As the number of agents increases from 1 worker to 3 and 6 workers, MINDcraft fails to complete any of the three builds. Meanwhile, CraftUtopia consistently succeeds and scales effectively: the build time significantly reduces as more workers are added (see the top right corner of Fig. 1). For example, building the NCPA with 6 workers is 1.5x faster than with 3 workers, and 3.9× faster than with 1 worker. In addition, we observe emergent behaviors similar to humans. Without any explicit instructions, workers discover how to use scaffolding to access elevated areas, and later remove it once it is no longer needed. Toward the end of a build, some workers naturally switch into a bystander, standing aside instead of hindering in ways that disrupt others. A full demonstration is available at: https://github.com/craftutopia-demo/CraftUtopia.

2 METHODOLOGY

CraftUtopia operates through two stages: design and build.

Design. A 2D architectural image cannot be built directly; it must first be converted into a Minecraft-compatible 3D model. In CraftUtopia, the designer achieves this in 3 steps: (i) reconstruct a 3D model from the image with TRELLIS.2; (ii) convert it into a block-based Minecraft blueprint via ObjToSchematic; and (iii) use LLMs to compile the 3D blueprint into a file that explicitly specifies block types and placements. Once created, the final blueprint is then passed to the project manager for downstream construction.

Build. To improve construction efficiency, we use two mechanisms: *hierarchical coordination* and *skill acquisition*. First, we organize agents into a hierarchical structure of "manager → foreman → worker". Given the blueprint, the manager splits the build into team-level subtasks over spatially disjoint regions for parallel execution. Each foreman then plans for its workers, who execute in Minecraft. Second, construction often repeats common routines (e.g., placing or removing blocks). Thus, we consider skill acquisition to distill recurring routines into reusable skills. Initially, the shared foremen skill library is empty, and foremen generate plans with LLMs. When a foreman detects frequently-repeated worker routines, it reports the pattern to the project manager, who codifies them into executable skills and shares them via the shared skill library. Foremen can then invoke these skill directly instead of re-planning with LLMs, which reduces planning latency and enables skill sharing across teams, and improve the speed over time.

3 CONCLUSION

Here, we introduce CraftUtopia, a novel LLM-based MAS for Minecraft that constructs 3D structures from a single 2D image. Through a Design-then-Build pipeline and two key mechanisms, CraftUtopia delivers various types of builds effectively and efficiently.

REFERENCES

[1] Saaket Agashe, Yue Fan, Anthony Reyna, and Xin Eric Wang. 2025. Llm-coordination: evaluating and analyzing multi-agent coordination abilities in large language models. In *Findings of the Association for Computational Linguistics: NAACL 2025*. 8038–8057.

[2] Qi Chai, Zhang Zheng, Junlong Ren, Deheng Ye, Zichuan Lin, and Hao Wang. 2025. Causalmace: Causality empowered multi-agents in minecraft cooperative tasks. *arXiv preprint arXiv:2508.18797* (2025).

[3] Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chi-Min Chan, Heyang Yu, Yaxi Lu, Yi-Hsin Hung, Chen Qian, et al. 2024. AgentVerse: Facilitating Multi-Agent Collaboration and Exploring Emergent Behaviors.. In *ICLR*.

[4] Yubo Dong, Xukun Zhu, Zhengzhe Pan, Linchao Zhu, and Yi Yang. 2024. Villageragent: A graph-based multi-agent framework for coordinating complex task dependencies in minecraft. *arXiv preprint arXiv:2406.05720* (2024).

[5] Ryusuke Fujisawa, Naohisa Nagaya, Shinya Okazaki, Ryota Sato, Yusuke Ikemoto, and Shigeto Dobata. 2015. Active modification of the environment by a robot with construction abilities. *ROBOMECH Journal* 2, 1 (2015), 9.

[6] Shiying Hu, Zengrong Huang, Chengpeng Hu, and Jialin Liu. 2024. 3d building generation in minecraft via large language models. In *2024 IEEE Conference on Games (CoG)*. IEEE, 1–4.

[7] Shuyue Hu, Siyue Ren, Yang Chen, Chunjiang Mu, Jinyi Liu, Zhiyao Cui, Yiqun Zhang, Hao Li, Dongzhan Zhou, Jia Xu, et al. 2025. Hands-on LLM-based Agents: A Tutorial for General Audiences. (2025).

[8] YICHEN JIANG, SUORONG YANG, SHENGJI TANG, SHENGHE ZHENG, and JIANJIAN CAO. 2025. A Comprehensive Survey of LLM-Driven Collective Intelligence: Past, Present, and Future. (2025).

[9] Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, et al. 2023. Agentbench: Evaluating llms as agents. *arXiv preprint arXiv:2308.03688* (2023).

[10] Chris Madge and Massimo Poesio. 2024. A llm benchmark based on the minecraft builder dialog agent task. *arXiv preprint arXiv:2407.12734* (2024).

[11] Nils Napp and Radhika Nagpal. 2014. Robotic construction of arbitrary shapes with amorphous materials. In *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 438–444.

[12] Kirstin H Petersen, Nils Napp, Robert Stuart-Smith, Daniela Rus, and Mirko Kovac. 2019. A review of collective robotic construction. *Science Robotics* 4, 28 (2019), eaau8479.

[13] Olivier Schipper, Yudi Zhang, Yali Du, Mykola Pechenizkiy, and Meng Fang. 2025. Pillagerbench: Benchmarking llm-based agents in competitive minecraft team environments. In *2025 IEEE Conference on Games (CoG)*. IEEE, 1–15.

[14] Haochen Sun, Shuwen Zhang, Lujie Niu, Lei Ren, Hao Xu, Hao Fu, Fangkun Zhao, Caixia Yuan, and Xiaojie Wang. 2025. Collab-Overcooked: Benchmarking and evaluating large language models as collaborative agents. *arXiv preprint arXiv:2502.20073* (2025).

[15] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2024. Voyager: An Open-Ended Embodied Agent with Large Language Models. *Transactions on Machine Learning Research* (2024). https://openreview.net/forum?id=ehfRiF0R3a

[16] Ziming Wei, Bingqian Lin, Zijian Jiao, Yunshuang Nie, Liang Ma, Yuecheng Liu, Yuzheng Zhuang, and Xiaodan Liang. 2025. MineAnyBuild: Benchmarking Spatial Planning for Open-world AI Agents. *arXiv preprint arXiv:2505.20148* (2025).

[17] Isadora White, Kolby Nottingham, Ayush Maniar, Max Robinson, Hansen Lillemark, Mehul Maheshwari, Lianhui Qin, and Prithviraj Ammanabrolu. 2025. Collaborating Action by Action: A Multi-agent LLM Framework for Embodied Reasoning. *arXiv preprint arXiv:2504.17950* (2025).

[18] Sipeng Zheng, jiazheng liu, Yicheng Feng, and Zongqing Lu. 2024. Steve-Eye: Equipping LLM-based Embodied Agents with Visual Perception in Open Worlds. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=NltzxpG0nz