

# HUBSIM: Simulating Human Behavior In Negotiation Settings

Nico Döring, Second Author<sup>2,3</sup>[1111–2222–3333–4444], Third Author<sup>3</sup>[2222–3333–4444–5555], and Fourth Author

<sup>1</sup> Princeton University, Princeton NJ 08544, USA

<sup>2</sup> Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany

`lncs@springer.com`

<http://www.springer.com/gp/computer-science/lncs>

<sup>3</sup> ABC Institute, Rupert-Karls-University Heidelberg, Heidelberg, Germany  
`{abc,lncs}@uni-heidelberg.de`

**Abstract.** The abstract should briefly summarize the contents of the paper in 150–250 words.

**Keywords:** First keyword · Second keyword · Another keyword.

## 1 Introduction

Add an introduction here.

## 2 Related Work

The rapid improvements of recent large language model capabilities prompted an increasing interest in exploring the potential for agent-based experimental studies and simulations in social science research. Park et al. [4] simulate and explore an entire human society and its behavioral interactions through an LLM-powered agent architecture which lets agents observe, plan and reflect. The framework of Qian et al. [5] shows how LLM agents exhibit efficient collaborative and self-corrective abilities in a software development scenario. Similarly, [2] conduct experiments on LLM-based research agents which perform classical machine learning research involving tasks such as data processing, architecture design and model training. Törnberg et al. [6] use LLM-based agents in different social media settings to simulate human behavior on a spectrum of toxic discourse and constructive interactions.

[4] uses gpt3.5-turbo [5] uses gpt3.5-turbo-16k [6] use gpt-3.5 [2] use gpt-4  
Negotiation Settings:

Evaluation of Agents: [1] evaluate chat-bison, claude-2, command, command-light, gpt-4, gpt-3.5 and (Llama2-70B-Chat) [7] evaluate chatbots based on gpt-3.5-turbo and GLMPro [3] AgentBench Evaluation Benchmark

Frameworks for developing agents: - Langchain Agents - AutoGen - CrewAI

## **3 Experimental Setup**

### **3.1 Research Design and Methodology**

Here goes general information, the approach and an explanation of our evaluation methods (quantitative and qualitative).

### **3.2 Negotiation Setup**

Here the setup for the negotiations (fixed-length and open-ended).

### **3.3 Interview Setup**

Here the setup for evaluation agent.

### **3.4 Parsing**

Here info about our parsing heuristic both for conversations and interviews.

## **4 Results**

### **4.1 Quantitative**

Quantitative Results.

### **4.2 Qualitative**

Qualitative Results.

## **5 Limitation**

Limitations here.

## **6 Conclusion**

Conclusion here.

## References

1. Davidson, T.R., Veselovsky, V., Josifoski, M., Peyrard, M., Bosselut, A., Kosinski, M., West, R.: Evaluating Language Model Agency through Negotiations (Jan 2024), <http://arxiv.org/abs/2401.04536>, arXiv:2401.04536 [cs]
2. Huang, Q., Vora, J., Liang, P., Leskovec, J.: Benchmarking Large Language Models As AI Research Agents (Oct 2023). <https://doi.org/10.48550/arXiv.2310.03302>, <http://arxiv.org/abs/2310.03302>, arXiv:2310.03302 [cs]
3. Liu, X., Yu, H., Zhang, H., Xu, Y., Lei, X., Lai, H., Gu, Y., Ding, H., Men, K., Yang, K., Zhang, S., Deng, X., Zeng, A., Du, Z., Zhang, C., Shen, S., Zhang, T., Su, Y., Sun, H., Huang, M., Dong, Y., Tang, J.: AgentBench: Evaluating LLMs as Agents (Oct 2023). <https://doi.org/10.48550/arXiv.2308.03688>, <http://arxiv.org/abs/2308.03688>, arXiv:2308.03688 [cs]
4. Park, J.S., O'Brien, J.C., Cai, C.J., Morris, M.R., Liang, P., Bernstein, M.S.: Generative Agents: Interactive Simulacra of Human Behavior (Aug 2023). <https://doi.org/10.48550/arXiv.2304.03442>, <http://arxiv.org/abs/2304.03442>, arXiv:2304.03442 [cs]
5. Qian, C., Cong, X., Liu, W., Yang, C., Chen, W., Su, Y., Dang, Y., Li, J., Xu, J., Li, D., Liu, Z., Sun, M.: Communicative Agents for Software Development (Aug 2023), <http://arxiv.org/abs/2307.07924>, arXiv:2307.07924 [cs] version: 3
6. Törnberg, P., Valeeva, D., Uitermark, J., Bail, C.: Simulating Social Media Using Large Language Models to Evaluate Alternative News Feed Algorithms (Oct 2023), <http://arxiv.org/abs/2310.05984>, arXiv:2310.05984 [cs]
7. Wang, Z.M., Peng, Z., Que, H., Liu, J., Zhou, W., Wu, Y., Guo, H., Gan, R., Ni, Z., Zhang, M., Zhang, Z., Ouyang, W., Xu, K., Chen, W., Fu, J., Peng, J.: RoleLLM: Benchmarking, Eliciting, and Enhancing Role-Playing Abilities of Large Language Models (Oct 2023), <http://arxiv.org/abs/2310.00746>, arXiv:2310.00746 [cs]