AI VILLAGE

Exploring AI interactions and dynamics in a simulated village

GROUP 1
311 - Advaith Arvind
320 - Arjun G Ravi
329 - Govind Sarath

Project guide: Dr. Subu Surendran

CONTENTS

- Aim
- Introduction
- Design
- Methodology
- Desired Impact
- Literature Survey
- Conclusion
- Reference

AIM

The aim of this project is to explore and analyze the behaviors, interactions, and emerging relationship dynamics among artificial intelligence (AI) agents within a simulated village environment. We also aim to learn about human societies through observing and analyzing AI agent interactions.

INTRODUCTION

Our goal involves crafting a virtual village using Pygame, placing full authority in the hands of autonomous AI agents. The agents will be able manipulate the world around it, which will inturn affect other AI agents. We intend to create a village where the AI agents coexist. We also intend to constantly monitor the relationship between the AI agents.

DESIGN

- Design a richly interactive virtual village utilizing Pygame
- Empower AI agents to independently manage and modify their surroundings
- Encourage AI agents to cooperatively inhabit shared spaces
- Continually monitor and assess 'relationship' amongst AI agents and the 'energy level' of the agents
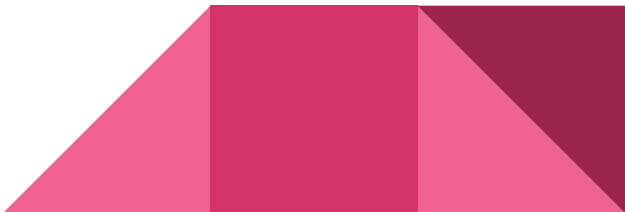
PEAS Analysis

- **Performance:** The 'energy' value for each AI agent. The 'resources' collected by each AI agent.
- **Environment:** The environment is the village simulated using pygame.
- **Actuators:** The environment is affected by the output of each (LLM) AI agent.
- **Sensors:** The prompt received by each LLM, acts as the sensor to observe the environment.

METHODOLOGY

1. A harmonious world where all AI agents interact with each other.
2. We introduce a limited resource, and incentivise the agents to collect the resources.
3. We create an election, where each AI agents try to impress other, to gain political support
4. We add a killer into the village. This might adversely affect the relationship between the agents
5. We introduce further chaos into the village, in the attempt to create further interesting observations

DESIRED IMPACTS

- Generate novel insights into AI behavior and social dynamics
- Advocate for conscientious AI deployment and collaboration
- Highlight pressing ethical concerns surrounding AI behavior and autonomy
- Uncover previously unidentified patterns and tendencies in AI interactions
- Validate existing theories related to multi-agent coordination and cooperation
- Study human societies through observing and analyzing AI agent interactions.

LITERATURE SURVEY

The primary research papers that aided the project are

- **Generative Agents: Interactive Simulacra of Human Behavior**
- Inner Monologue: Embodied Reasoning through Planning with Language models
- Improving Factuality and Reasoning in Models through Multiagent Debate
- Character-LLM: A Trainable Agent for Role-Playing
- Chain of Thought Prompting Elicits Reasoning in Large Language
- AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions

1. Generative Agents: Interactive Simulacra of Human Behavior

- This is the primary paper that introduced the idea of AI agent simulation with large language models.
- The paper creates a virtual AI village where AI agents live together in harmony.
- The paper talks about how the AI agents are able to organise a party, by communicating with fellow agents.

2. Inner Monologue: Embodied Reasoning through Planning with Language models

- This paper focuses on giving LLMs an 'Inner Monologue'.
- It is to investigate to what extent LLMs used in embodied contexts can reason over sources of feedback provided through natural language when additional training is not provided.
- The inner monologue allows the LLMs to more richly process and plan robotic control scenarios.

3. Improving Factuality and Reasoning in Models through Multiagent Debate

- This paper presents a complementary approach to improve language responses where multiple language model instances propose and debate their individual responses and reasoning processes over multiple rounds to arrive at a common final answer.
- The findings indicate that this approach significantly enhances mathematical and strategic reasoning across a number of tasks.

4. Character-LLM: A Trainable Agent for Role-Playing

- The paper addresses an intriguing subject of how characters can be simulated using LMs.
- The authors devise a novel methodology and playground for making progress on this that allows them to train agents that they then evaluate for their memories/experiences.
- The paper creates a new LLM (Character LLM) which is fine tuned to behave just like a human being.

5. Chain of Thought Prompting Elicits Reasoning in Large Language

- Chain of thought prompting enables models to generate intermediate reasoning steps to help solve multi-step arithmetic, commonsense, and symbolic reasoning tasks.
- It shows us how much simple logical analysis of LLMs can make to the output.
- We are planning to implement such a chain of thought prompting into our AI agents.

6. AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions

- This paper studies the impact of artificial intelligence on human-machine interaction, particularly through conversational AI driven by machine learning, deep learning, and natural language processing.
- Furthermore, modern conversational agents are designed to simulate human behavior by incorporating emotional, sentimental, and affective nuances.
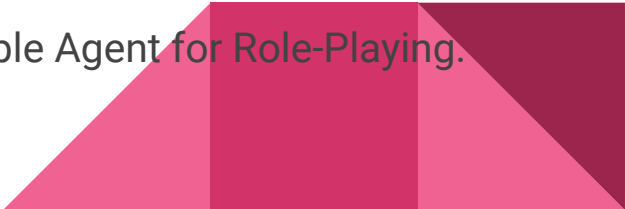
CONCLUSION

To conclude, our AI Village project demonstrates great promise in advancing understanding of AI behavior and social dynamics within intricate environments. Its emphasis on responsible AI integration, collaborative coexistence, and continuous relationship assessment offers unique opportunities for exploring novel implications and consequences arising from AI presence among us. Ultimately, we hope to inspire informed discussions regarding AI ethics, governance, and societal impact.

REFERENCES

[1] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In ACM, New York, NY, USA.

[2] Huang, W., Xia, F., Xiao, T., Chan, H., Liang, J., Florence, P.R., Zeng, A., Tompson, J., Mordatch, I., Chebotar, Y., Sermanet, P., Brown, N., Jackson, T., Luu, L., Levine, S., Hausman, K., & Ichter, B. (2022). Inner Monologue: Embodied Reasoning through Planning with Language Models. Conference on Robot Learning.

[3] Mialon, G., Dessì, R., Lomeli, M., Nalmpantis, C., Pasunuru, R., Raileanu, R.,Rozière, B., Schick, T., Dwivedi-Yu, J., Celikyilmaz, A., Grave, E., LeCun, Y., & Scialom, T. (2023). Augmented Language Models: a Survey. ArXiv, abs/2302.07842.

[4] Park, J.S., Popowski, L., Cai, C.J., Morris, M.R., Liang, P., & Bernstein, M.S. (2022). Social Simulacra: Creating Populated Prototypes for Social Computing Systems. Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology.

[5] Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J.,Chen, X., Lin, Y., Zhao, W.X., Wei, Z., & Wen, J. (2023). A Survey on Large Language Model based Autonomous Agents. ArXiv, abs/2308.11432.

[6] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022).ReAct: Synergizing Reasoning and Acting in language Models. ArXiv,abs/2210.03629.

[7] Anil, Gemini Team Google Rohan et al. "Gemini: A Family of Highly Capable Multimodal Models." ArXiv abs/2312.11805 (2023): n. pag.

[8] Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T.L., Cao, Y., & Narasimhan, K. (2023). Tree of Thoughts: Deliberate Problem Solving with Large Language Models. ArXiv, abs/2305.10601.

[9] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E.H., Xia, F., Le, Q., & Zhou, D. (2022). Chain of Thought Prompting Elicits Reasoning in Large Language Models. ArXiv, abs/2201.11903.

[10] Shao, Y., Li, L., Dai, J., & Qiu, X. (2023). Character-LLM: A Trainable Agent for Role-Playing. ArXiv, abs/2310.10158.

[11] Du, Y., Li, S., Torralba, A., Tenenbaum, J.B., & Mordatch, I. (2023). Improving Factuality and Reasoning in Language Models through Multiagent Debate. ArXiv, abs/2305.14325.

[12] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, S. Mishra and A. Abraham, "AI-Based Conversational Agents: A Scoping Review From Technologies to Future directions," in IEEE Access, vol. 10, pp. 92337-92356, 2022, doi: 10.1109/ACCESS.2022.3201144.

Thank you