GPT-NPC: Enhancing NPC Human-Likeness and Autonomy in Video Games

Author: David Ogunlesi and Xiaoyang Wang
Published in :European Conference on Artificial Intelligence (ECAI) 2024 Workshop on "eXtended Reality & Intelligent Agents" (XRIA 24)

YAHAGI Takuya

Introduction

- Non-player Characters (NPCs) are one of the factor of deciding player's immersion.
- Conventional NPCs only speak what programmed in advance, and hinder immersion.

- The greater improvement of NLP. such as Transformer, or LLM, has a possibility to access this issue.
- However these conventional methods has still some problems.

Conventional Methods

- Rule-based system for NPC.
  - Finite State Machine (FSM)
  - Behaviour Tree
  - Goal-Oriented Action Planning (GOAP)

- Limitation
  - Conversation is limited in the range of settings in advance.
  - Scalability problem because it is set manually.

Conventional Methods

- NPC with LLM can spread flexibility of conversation and improve human likeness.

- Limitation:
  - The limitation of context window length.
  - The limitation of single LLM, and lack of human aspect.
  - The lack of research which incorporating LLM into the digital space.

Purpose

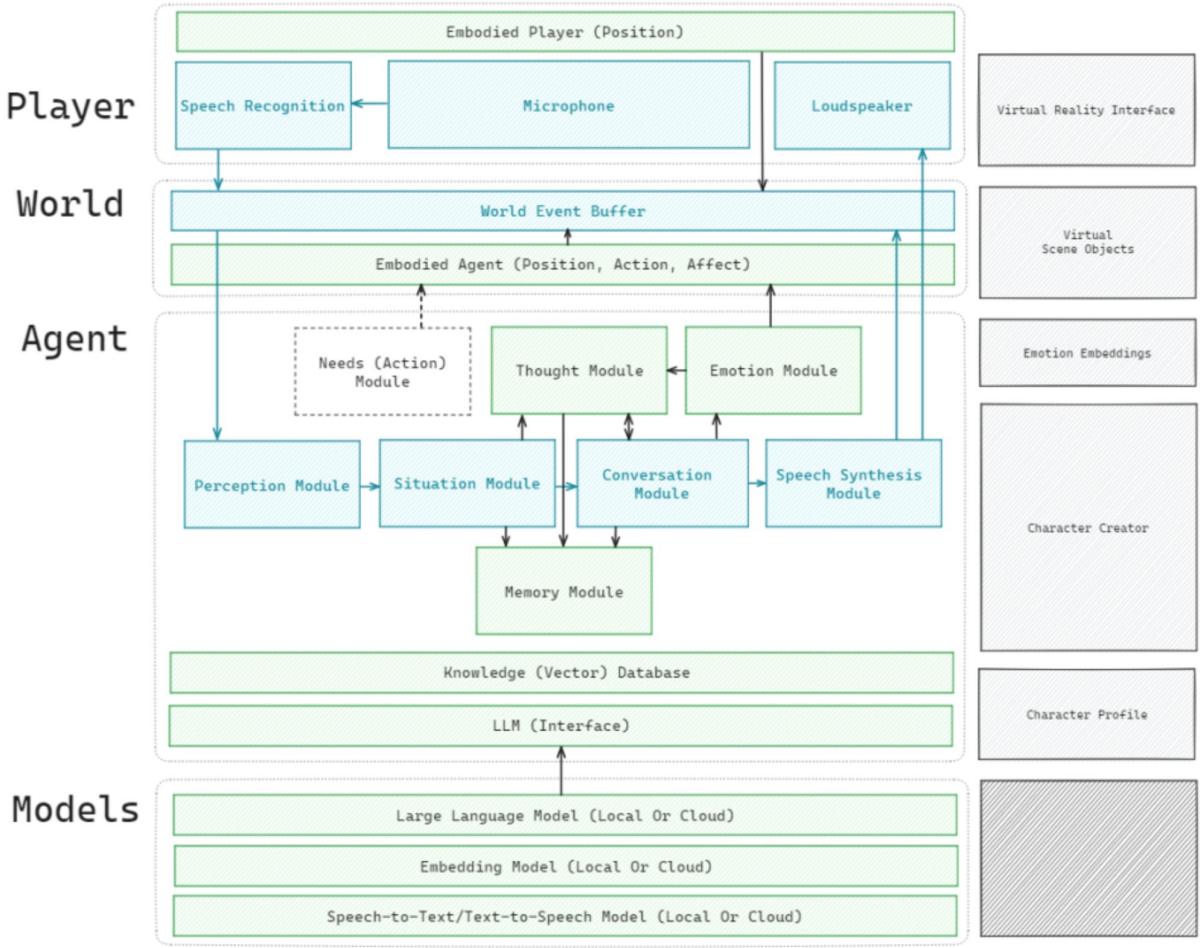
- The researchers suggested not single LLM approach, but Modular Asynchronous Approach.
- In simple terms, they suggested GPT-Team to make NPC more human-like with consistent text generation.

- They investigated this approach in the Virtual Reality game, they tried to make that LLMs work as a materialized agent, it perceive physical body, any sense, and make decision and interact with environment.

Methods

- In their suggested system, there are 7 modules:
  - Perception : Recognize surrounding environment, and convert it into internal expression.
  - Situation : Summarize all perceived events, then understand what is happening now comprehensively
  - Conversation : Based on internal state, decide what speak next, to make consistent and attractive text.
  - Speech Synthesis : Just do a task of text-to-speech.

  - Memory : Memorize history of important event or conversation, and pile it as memory.
  - Thought : Generate “thoughts” autonomously, gives NPC the ability to reflect on itself.
  - Emotion : Predict the player’s emotion, and generate emotional response.


Player
World
Agent
Models
Embodied Player (Position)
Speech Recognition
Microphone
Loudspeaker
World Event Buffer
Embodied Agent (Position, Action, Affect)
Needs (Action) Module
Thought Module
Emotion Module
Perception Module
Situation Module
Conversation Module
Speech Synthesis Module
Memory Module
Knowledge (Vector) Database
LLM (Interface)
Large Language Model (Local Or Cloud)
Embedding Model (Local Or Cloud)
Speech-to-Text/Text-to-Speech Model (Local Or Cloud)
Virtual Reality Interface
Virtual Scene Objects
Emotion Embeddings
Character Creator
Character Profile

Experiment

- Evaluated the conversation with NPC in Virtual Reality (VR) environment.
  - Participants: 15 students of University, from 18 years-old to 26 years-old.
  - Environment: The classroom with a NPC, Morrowii, in VR.

Experiment

- Each participant speaks with Morrowii in 3 patterns of settings, without thought and emotion module, with only thoughts module, and with both modules.

- After each conversation, they evaluate human-likeness with 6 criteria:
  - Coherence : is the GPT agent being logical and consistent?
  - Memory : how well does the GPT agent remember the context?
  - Awareness : is the agent knowledgable and aware of the situation?
  - Emotions : can the GPT agent understand emotions?
  - Autonomy : can the agent behave autonomously?
  - Latency : how do you score the experienced latency in this interaction?

Evaluation

- After evaluation for each criterion, they take the weighted sum based on the degree of importance.
- Bigger weight means Bigger importance.

- Then, calculated value is the coherence score of human-likeness for each setting.

Success criterion	Symbol	Weight
Coherence	S_1	8
Memory	S_2	6
Awareness	S_3	7
Emotions	S_4	5
Autonomy	S_5	2
Latency	S_6	10

Results

		No Thoughts, No Emotions	Thoughts, No Emotions	Thoughts, Emotions
Success Criteron	Coherence	0.8	0.8	0.8
	Memory	0.6	0.6	0.6
	Awareness	0.5	0.65	0.65
	Emotions	0.6	0.7	0.85
	Autonomy	0.71	0.83	0.83
	Latency	0.49	0.37	0.35
Evaluation Score		**0.6**	**0.616**	**0.63**

- Overall, with thoughts and emotions, the NPC's human-likeness were improved
- Especially, emotions module enable NPC to express their emotions more human-likely.
- Latency is the biggest problem, it takes 2-4 seconds for latency.
  -> Cause to decrease reliability of NPC.

Analysis (There are no discussion section…)

- Thoughts and emotions module is important to improve human-likeness of NPC.
- Especially, emotions module NPC express more abundantly, the score of emotion got higher.
  -> Player feel NPC as more “living”.

- However, latency is one of the biggest problems to decrease NPC’s reliability.
  -> It shows that the what important for human-likeness is not deepness but timing of conversation.
- The main cause of this latency is the dependency with API calling.
  They should do technical improvement such that use of local AI model, or the hardware that can process models with high speed.

Conclusion

- This research shows the possibility of the GPT-NPC framework that can improve autonomy and human-likeness.
- They thought that combination of each module can make more attractive and complex character.
- However, the latency is the biggest problem for the practical use.
- In the future, it is needed that the high speed local model, more efficient way to process, etc for improving these systems.

Thank you for listening!