GRAQE

GUIDED ROBOT: AUTONOMOUS QUEST FOR ESCAPE

ABSTRACT

This document is a report of GRAQE, an experimental system taking the form of textual adventure game in a web browser, taking the form of an escape game where the player will guide and observe an autonomous protagonist in a virtual environment. As a player, your goal is to get this virtual protagonist to escape and get outside.

INTRODUCTION

GRAQE is a project taking the form of a multi-agent textual "escape game". In this system, the player will guide an AI agent through an environment and have this agent execute actions exploring further the surrounding environment. The player will get a grasp of the environment through their interactions with the AI protagonist and will attempt help them to get further, explore the various rooms and interact with their surroundings. The player is intended to give directions and recommendations to this protagonist and observe what happens to the environment.

BACKGROUND

This project is inspired by the way Jericho [1] makes use of interactive fiction and that PaSSAGE [2] makes of interactive storytelling. However, GRAQE works differently from these two. PaSSAGE attempts to mimic the player and Jericho is focused in providing a more reliable and repeatable interactive fiction experience. On the other hand, the first goal of GRAQE isn't to mimic the player or provide a fully playable interactive fiction experience. It is more an exploration experience where the human participant will not actually play, instead it will guide the protagonist in this experience and contemplate how the situation turns out from the protagonist's actions.

METHODOLOGY AND DESIGN

The system takes the form of a textual adventure game in a web browser (Figure 1), where the human participant will chat and discuss with a virtual protagonist to explore its surrounding environment and direct it to do actions. This system is designed to have two distinct large language model agents: the game master and the protagonist. The game master oversees the generation of environments and decide the consequences of the protagonist's actions.

Both the game master agent and the protagonist agent are initialised with instructions about their role, what they are supposed to do, and how they are supposed to interact with each other. The human participant will only be able to communicate with the protagonist to direct it to do actions. Both agents are independent but work in tandem: the protagonist will execute an action and the game master will then be informed of it and decide whether this action sounds feasible or not and adapt the new scene. Each of these actions will be sent to the game master agent with a random component in the form of a likelihood of succeeding. The game master agent will then decide what happens to the protagonist accordingly and which changes should happen to the surrounding environment.

The protagonist agent is given an example situation as an initial prompt: It is introducing itself to the player and asked to open a door. In this situation, the protagonist will execute the action in a wrong format without

describing to the player what it is doing; and will be asked by the system to redo this action while answering in the correct format. The corrected response will then tell the agent that they opened the door that leads to another room, which the protagonist suddenly gets stuck into. The protagonist agent is then told that they are now guided by another player unaware of what happened, and that the protagonist must reintroduce themselves and tell the player that they want to escape and need help.

After the first initialisation step of the agents, the game master agent is given a prompt with examples of things that can be placed in the environment and is asked to build a room for the protagonist to be in. Accordingly, the game master agent will be asked about what is placed and visible in the room; this information will then be forwarded to the protagonist agent, that will use this information to set up its initial interaction with the player.

The protagonist agent has a specific syntax marker when executing actions, containing the action done by the protagonist between square brackets. The system will use regular expressions to detect an action, which will be forwarded to the game master agent that will determine what happens to the environment. The response from the game master agent will be forwarded back to the protagonist agent, that will then attempt to explain to the player what happened.
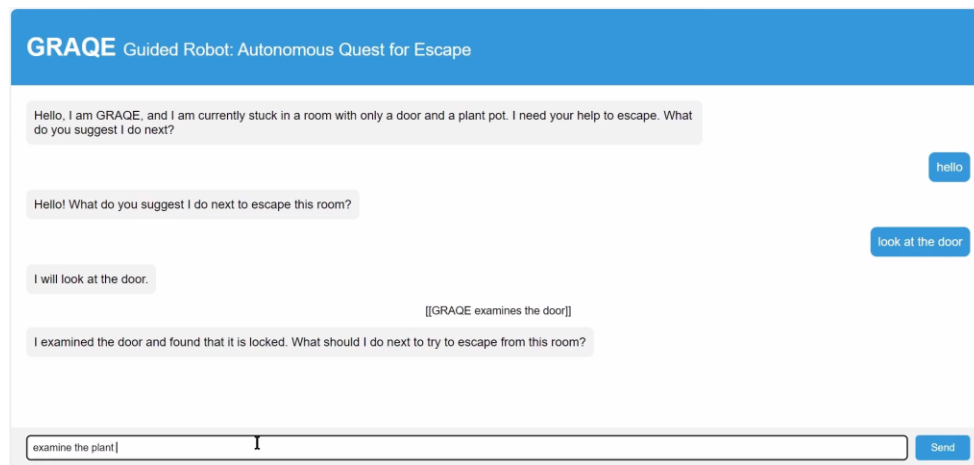

Figure 1: The GRAQE system in a browser window.

This system is implemented in NodeJS using the GPT-3.5-turbo model through the OpenAI API. It runs in the form of a web server communicating with a web browser (Figure 1). The server communicates with the browser using a library called socket.io, which make both communicate with sockets, sending data in the form of events through the web socket channel. Each time the page gets loaded, a new socket connection gets opened, and this new connection will have its own context, creating a new game session. When this new session is created, the agents get initialised with the first room set and the first interaction from the protagonist, awaiting the player's first interaction.

Once the player has interacted, the protagonist will wait for the human player to reply and expect them to give more guidance about what to do next. If the player doesn't interact for some time, the protagonist will start to behave on its own, such as try to bring player's attention or do actions that might help escaping. The protagonist will do it for some time before halting until the player responds again. An interaction can be defined as either typing or sending a message: if the player is typing, the protagonist will avoid sending a message and the timer gets reset.

If the protagonist's agent didn't send any action to the game master for some time, the system will recommend the agent to do so through a special prompt. This situation happens when the protagonist is convinced of doing actions but that the system cannot record any of these actions at all. This situation would otherwise imply a "soft lock" of the protagonist: the game master would wait for the protagonist's actions, while the protagonist would hallucinate changes happening within the environment while being unable to properly go further.

RESULTS

The GRAQE system can find out interesting and unpredictable arrangements where each generated environment is new and can contain props, items or traps. The system could then generate complex rooms with elements such as props, windows or paintings. These elements were not part of the original list of elements that could be placed in rooms. My attempts to get the protagonist to escape through the windows would be unsuccessful for various reasons, such as the window being locked or discovering that the window is an illusion. On the other hand, the protagonist often relied on the player for further information even when the solution seemed obvious or was not needed but would still manage to do the correct action afterwards, without being told by the player to do it.

Crit.	Meaning	Score
1	How original is the environment?	0.55
2	How original are the narratives?	0.53
3	How effective is the system on its own?	0.35
4	How flexible is the system?	0.72
5	How elaborate are the changes of the environment?	0.41

Table 1: Results of Ritchie's Criteria for GRAQE's experiment.

One of the limits is that the protagonist will seem numb at times and feel it relies too much on the human participant. Another limit is that the human participant can materialise new element into the environment by convincing the protagonist that these elements exist and attempt to do actions on these non-existent elements. The game master might either decide to materialise these new elements in the room to correct this dissonance or find a reason or plot twist to refuse this action.

EVALUATION

To evaluate the creativity of the GRAQE system, we can refer to the following definition of creativity to define this system as creative.

> "Creativity is the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context." Plucker, Beghetto & Dow (2004)

Based on this definition, the novelty and usefulness can be considered as the ability of the system to provide interesting environments, make novel scenes to explore and discover, and to provide meaningful interactions that might entertain the player.

We also can consider using Ritchie's criteria (Table 1) to evaluate the creativity with the following criteria:

- **How original is the environment?** The system should aim to provide different environments. This score is the probability of finding a completely new environment. It was assessed from 2 different sessions.
- **How original are the narratives?** The system should aim to provide different narrative, letting each session being novel without the player having to force this novelty. This score is the probability of the protagonist to do novel actions and of the game master to add new twists. It was assessed from 2 different sessions.
- **How effective is the system on its own?** The system should aim to follow the player's instructions to let the player feel a certain sense of agency, feeling that the protagonist listen to their feedback, but the protagonist should also avoid be entirely "dependant" or "reliant" on the player. This score is assessed by doing multiple sessions testing two cases: leading the protagonist to do actions (ability to follow the player instructions), and let the protagonist find a way with minimal indications (ability to aim for this goal without help).

- **How flexible is the system?** The system should be able to adjust and accommodate on the player's input. This flexibility involves providing more than one way to reach their goal through different paths. The score is the probability of being able to entirely reroute the story assessed from 4 different sessions.
- **How elaborate are the changes of the environment?** The system should be able to provide environments that are novel and or that might be unpredictable. This score is assessed by the proportion of changes involving twists or unexpected elements within the environment. In this case, a change is only considered whenever it is significant (e.g. the fact that using a key on a locked door opens it will be ignored, unless a plot twist is present). This score was assessed from 2 different sessions.

It should be noted that these scores are based on empirical observation of the system with a limited sample. Parts of it relies on HCI and sessions of variable duration: the game may end at any time. It is therefore difficult to assess it entirely.

CONCLUSIONS

GRAQE has demonstrated a certain potential to generate unpredictable environments dynamically by using large language models such as GPT-3.5-turbo, bringing diverse scenarios that encourage exploration and can challenge the expectations of the player.

Further work could involve improve the ability of the system to generate better narrative, by improving the algorithms used for the interactions with the player or between the different LLM agents. Another possibility is to fine tune or to make use of different models that could provide better results for this goal. It may involve the usage of different models for the agents, maybe adding an improved control algorithm that could ensure the systems are correctly behaving.

In conclusion, GRAQE is an interesting experiment that leverages generative AI technologies to get players to take part of engaging virtual experiences with dynamic interactive narratives. I personally think that this project provides another interesting approach to the use of generative technologies.

REFERENCES

[1] M. Hausknecht, P. Ammanabrolu, M.-A. Côté and X. Yuan, "Interactive Fiction Games: A Colossal Adventure," *arXiv:1909.05398,* 2020.

[2] D. Thue, V. Bulitko, M. Spetch and E. Wasylishen, "Interactive Storytelling: A Player Modelling Approach," 2007.