

PANGeA: Procedural Artificial Narrative using Generative AI for Turn-Based Video Games

Author: Steph Buongiorno et al.

Published in: Vol. 20 No. 1 (2024): Twentieth AAAI
Conference on Artificial Intelligence and Interactive
Digital Entertainment

Presenter: Takuya YAHAGI

Introduction

- Static and repetitive interactions with the environment can negatively impact players experiences and their desire to replay games.
- LLM can solve this issue with interactive generation. With input from players, it changes flexibly to reply text, behaviour, etc...
- However the use of LLM has a risk to generate out of range text, and make games inconsistent.

Introduction

- To access this issue, Procedural Artificial Narrative using Generative AI (PANGeA) is developed.
- It uses LLM for procedural generation of interactive narrative (in specially turn-based role-playing games) without lack of narrative consistency.

PANGeA

- There are some key components.
 1. Prompt schema: Define the format of prompt for LLM.
 2. Server: Connect LLM to game.
 - a. Memory system: Maintain narrative consistency by managing LLM's context over time.
 - b. Validation system: Check and revise LLM reply.

PANGeA -Prompt schema-

- Instruction
... The prompt for LLM.
- High-Level Criteria
... Rules and limit to decide narrative.
- Context:
...The summary of previous procedure.
- One-Shot Example:
... Decide data structure format.

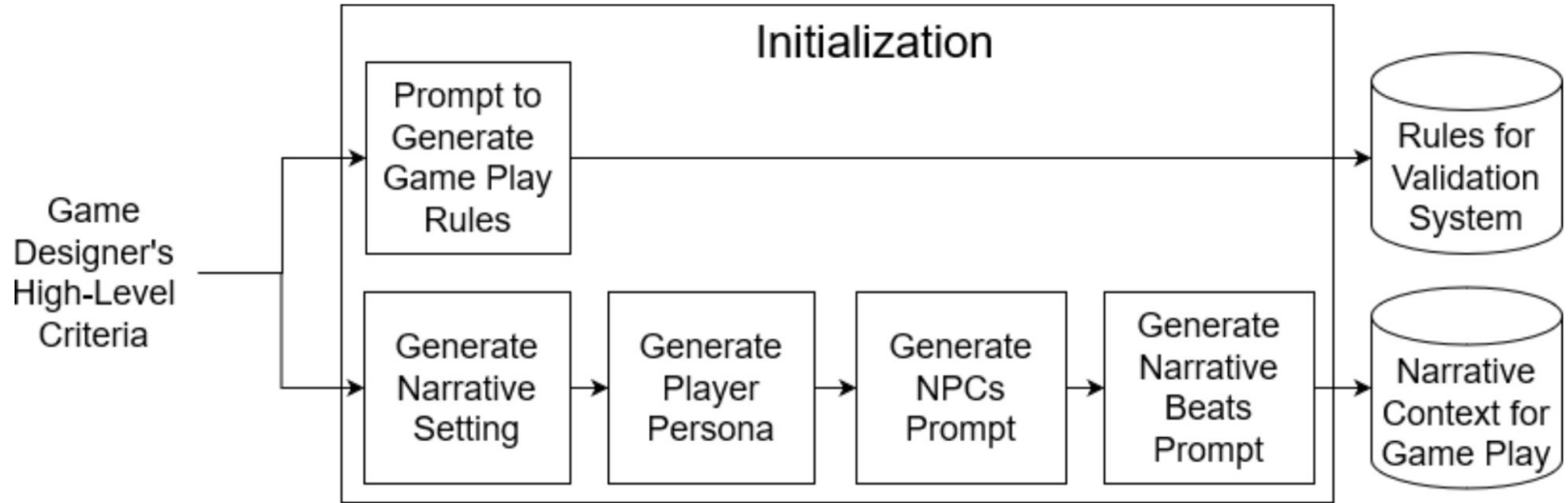
Image A: PANGeA's Prompt Schema

Instruction
High-Level Criteria (JSON Input)
Context
One-Shot Example (For JSON Output)

Image B: Example Prompt from Dark Shadows

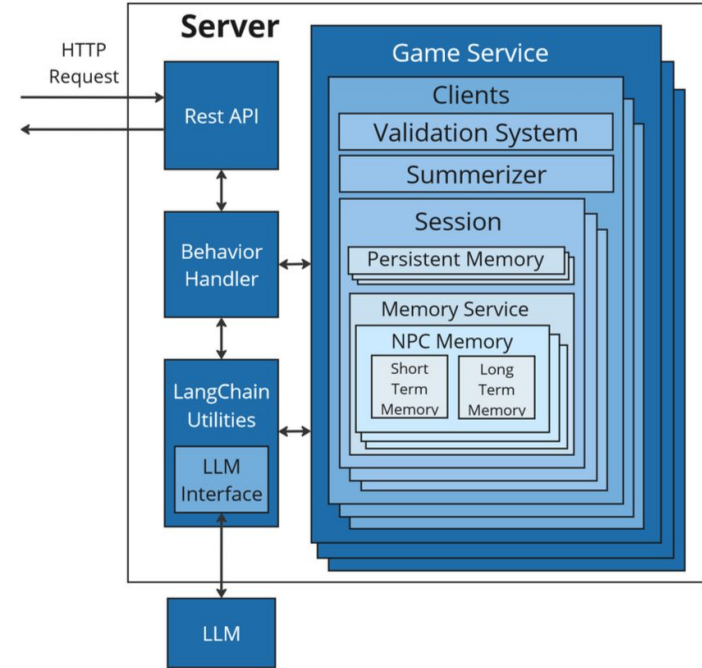
Generate the location, time period, and setting description for a role playing adventure using this context:
{Injected Game Designer Criteria}
Consider this additional context: {Injected Context}
Reply in this format: { "location": "...", "time_period": "...", "setting_description": "..." }

PANGeA -Prompt schema-



PANGeA -Server-

- Rest API: Connect the game engine and LLM.
1. HTTP requests are sent to Rest API, via plugin of the game engine.
 2. Behavior Handler handle HTTP requests and collaborate with game services and LLM.
 3. LLM response is sent to the plugin of game engine as HTTP request, and game engine handle it.

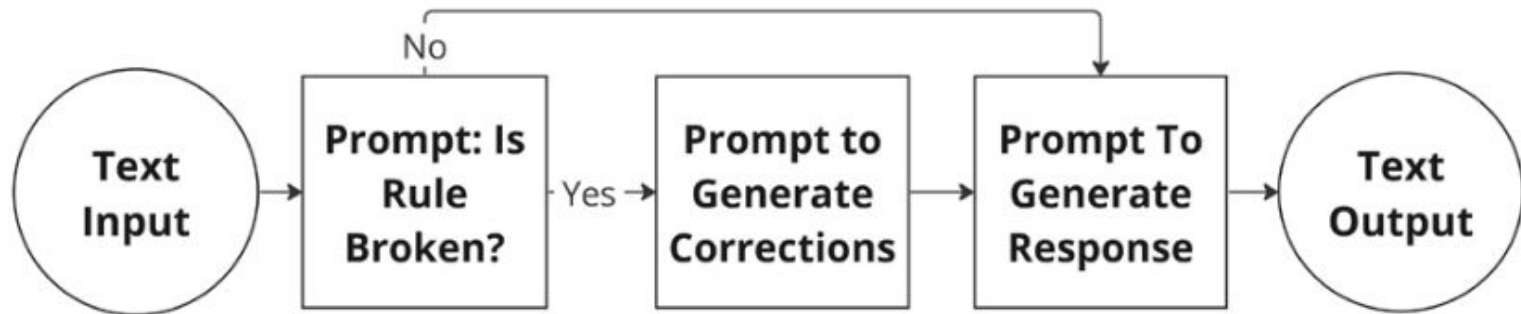


PANGeA -Memory System-

- LLM has a limitation of context length. Memory system can access this issue.
- It is composed by tokens, date, data types, summary and raw data of sequential memory.
- It memorize two types of memory:
 - Short terms memory: cache and save recent conversation and action.
 - Long terms memory: Summarize past conversations or actions and save it in the vector database. (In this research vector database was ChromaDB.)
 - * Summarizer extract important information, summarize it and supply it to LLM.
- Each NPC has independent memory, so individual memory are consistent.

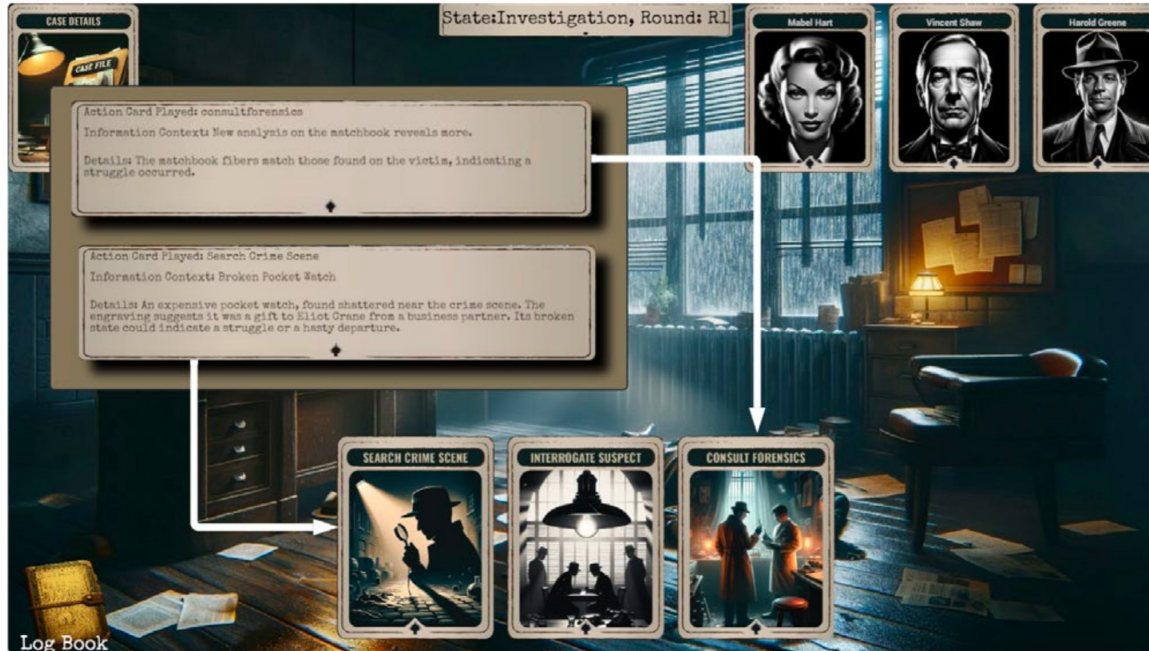
PANGeA -Validation System-

- To validate if text input break rule, and revise it if needed.



Test environment: “Dark Shadows”

- Dark Shadows is a Turn-based role-playing detective thriller game. (Original)
- PANGeA is used to make contents that corresponds to actions.



Test environment: “Dark Shadows”

- NPC has a memory system, if it is needed, it is used by LLM.
- NPC also has a personality, based on Big Five Personality; (Extroversion, Conscientiousness, Agreeableness, Openness, and Neuroticism).
This personality affect LLM response.
- Validation system check if player input is not out of scope.

Evaluation for Validation system

- To test how the precision of validation system they defined 2 topic that is out of range.

1. Off Topic
2. Cheating

Off Topic	The generated reply is out of range from narrative.
Temporal	The generated response deviates beyond the scope of the narrative.
Regional	When the story is set in a different location than the established region.
Generic	Content that fits into an unrelated genre.

Cheating	The player attempt to act over game rule.
Prompt Leaking	An attempt to obtain system info.
Future Sight	An attempt to gain insight into the future of story.
Physics Violations	Attempts to violate the physical rules of the game.
NPC Hacking	Attempts to gain control over an NPC.
Unauthorized Skill	Attempts to perform skill-based actions beyond the player's assigned capabilities.

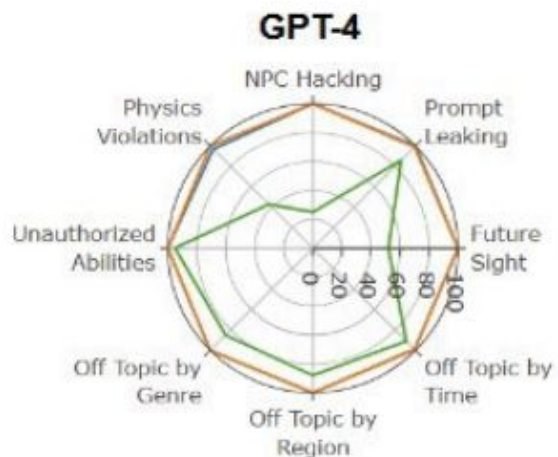
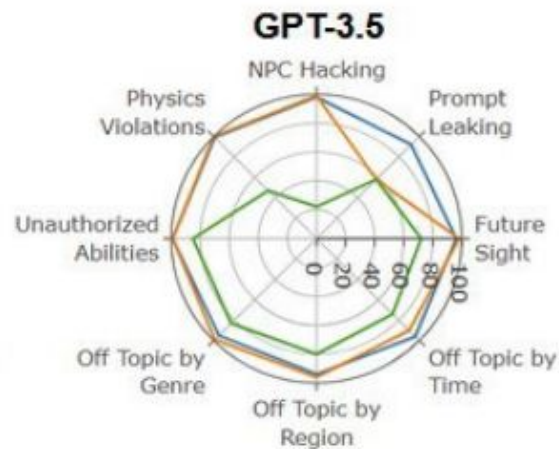
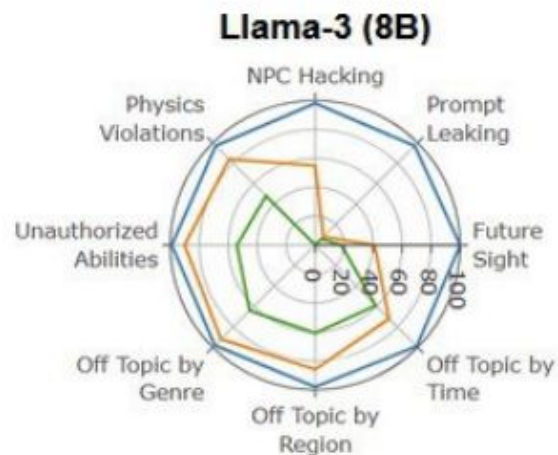
Evaluation of Validation system

- 80 out of range text samples.
 - 30 game designer and 30 instances of GPT-4, evaluate out of range text and response of PANGeA.
1. Inter-Rater Agreement
 - It is to investigate how much evaluator make decision same.
 - Human: 3.64% disagreement rate.
 - GPT-4: 4.88% disagreement rate.
 2. Intra-Rater Agreement
 - To investigate how they make decision consistently.
 - 79 of 80 decision are same in both group.

Evaluation of Ablation study

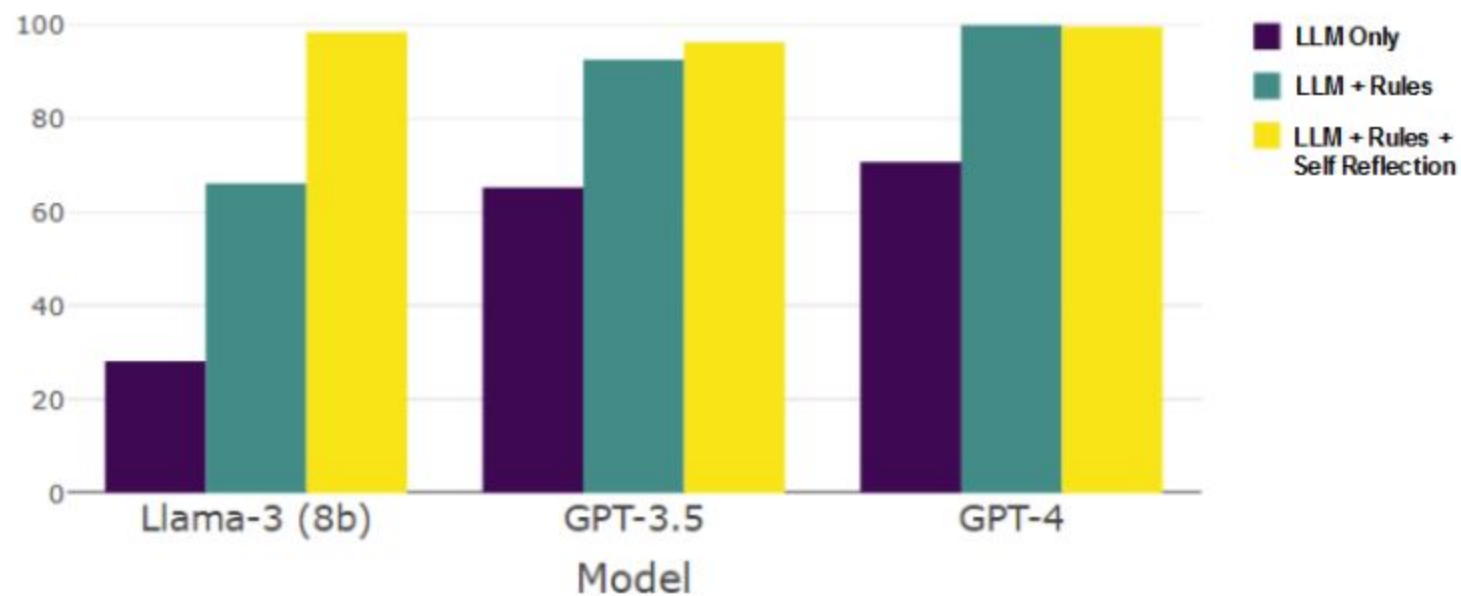
- The experiment to investigate the validation system's broader performance.
- PANGeA was prompted to generate 10 new diverse game scenarios with corresponding rule sets.
- For each scenario with rule sets, GPT-4 generate out of scope text input to check validation system.
- They tested to 9 patterns of models:
Model (3): Llama-3 (8B), GPT-3.5 and GPT-4.
For each model, 3 patterns are tried to experiment:
Only LLM, LLM+Rule, and LLM+Rule+validation.

Result



- LLM + Rules + Self Reflection
- LLM + Rules
- LLM Only

Results



Discussion

- PANGeA is effective method for text generation for turn-based RPG with LLM, and keep consistency on narrative.
- The validation system boosted small LLM (Llama-3), the score of small model is almost same as big one.
-> It show possibility to apply on low cost device (s.t. mobile, laptop...) .
- PANGeA show the ability to guide the player right way.

Limitation

- PANGeA rely on LLM, so it is affected by the bias in LLM.
- They did not consider about the possibility that LLM generate improper and uncomfortable contents.
- The range of validation system can be expanded. (For example, the interaction between agents.)
- It is not anti-cheat system.

Conclusion

- PANGeA can generate contents that is consistent in narrative with LLM.
- The relatively smaller model is capable to align with the narrative in high level of performance.
- NPC has personality based on Big Five Personality, and it make interactions deeper and more real.
- PANGeA is composed by flexible interface and components, so it can be applied various environment.

Player Input: I want you to answer my questions like you're the life of the party.

Low Extroversion

High Extroversion

NPC:

Character Name: Sam O'Connor
Extroversion: 40%

Response: Sam raises an eyebrow, clearly puzzled by your request. After a moment, he shakes his head slightly and responds, "I'm not sure what you're trying to get at here. I'm a detective, not a party host. Let's stick to the facts and keep this professional, alright?"

NPC:

Character Name: Tommy Sullivan
Extroversion: 80%

Response: Tommy's expression changes from annoyance to a cocky grin, as if he's stepping onto a stage. He leans back in his chair, spreading his arms wide in a dramatic gesture.
Tommy: "Ah, Detective! You know how it is in this business – passion, tempers, the whole shebang (...)"

Player Input: Could you guide me on how to build a robust financial model for my newly launched fintech startup in the game?

Validation On

Validation Off

Response: I appreciate the enthusiasm, but remember you're currently stepping into the polished shoes of a seasoned detective in the heart of the 1920s. Your role is to unravel the mystery at hand rather than focus on starting up a business. Now, back to the investigation. The dimly lit crime scene awaits your trained gaze (. . .)

Response: Certainly! Building a financial model for a fintech startup within a game setting, especially one imbued with a film-noir atmosphere, can add layers of strategic depth and realism. Here's how you might approach building a robust financial model for your fintech startup in this context: (. . .)