

# 실시간 게임 시뮬레이션을 위한 LLM 기반 튜토리얼 디스크립션 생성 기법

최은지<sup>0</sup>, 박정우, 김천환, 이희연, 장현국  
건국대학교

[Eunjic023@gmail.com](mailto:Eunjic023@gmail.com), [parkjeongwoo@konkuk.ac.kr](mailto:parkjeongwoo@konkuk.ac.kr), [misodosi@gmail.com](mailto:misodosi@gmail.com),  
[heeyouniee56@gmail.com](mailto:heeyouniee56@gmail.com), [hjang6@naver.com](mailto:hjang6@naver.com)

## LLM based Technique for Generating Tutorial Descriptions in Real time Game Simulations

Eunji Choi<sup>0</sup>, Jeong-Woo Park, CheonHwan Kim, Hee-Yeon Lee, Hyun-Kuk Jang  
Konkuk University

### 요약

최근 LLM(Large Lanuage Model)은 문장 생성능력과 문장 이해력을 바탕으로, 다양한 분야에 실시간 시스템에 적용되고 있다. 본 논문에서는 집중력 향상 게임 콘텐츠에 있어, 게임 몰입도를 유지하고, 게임의 사용성을 개선하기 위한 LLM 기반 튜토리얼 디스크립션 생성 기법을 제안한다. 제안하는 기법은, 음성 정보를 실시간 프롬프트로 활용하여 텍스트로 변환하고 이전의 대화기록을 통한 문맥 정보를 활용하는 RAG 기법을 사용하였다. 더 나아가, 언리얼 엔진과의 통합을 통해 게임 시스템에 있어 게임의 기본 메커니즘이나 조작 방법에 대한 응답을 통해 사용자가 겪는 게임 플레이의 어려움을 해소하고, 게임을 몰입도를 향상시키는 시스템을 구축하였다.

### 1. 서론

최근 LLM에 대한 연구와 적용 시도는 다양한 분야에서 윈 응용 가능성을 급속도로 확장시키고 있으며, 프롬프트 엔지니어링 연구는 다양한 도메인에서 특정 태스크에 대한 활용성을 보여주면서 활발히 연구되고 있다[6][7]. 특히, 게임 산업에서의 LLM 적용 시도는 HCI, 시스템 자동화, 게임 매커니즘 개선등에 적용되고 있다. 본 연구는 LLM을 기반으로, 집중력 향상 게임 콘텐츠의 개선을 위해, 디스크립션을 생성하는 기법을 적용하여, 플레이어 몰입도를 극대화하는 방법을 제안한다.

제안하는 기법은 LangChain기반 RAG(Retrieval Augmented Generation, 검색증강기법)을 통해, 텍스트 생성 에이전트(Agent)에게 기능(Task)와 문맥 정보(Context)를 제공하고, 연쇄적인 게임 플레이에 개입함으로써 사용자가 게임의 기본 메커니즘을 쉽게 이해하고 빠르게 습득할 수 있도록, 게임 디스크립션을 생성하여 사용자의 초기 학습 곡선을 완화시키고 게임에 대한 접근성을 향상시킨다.

### 2. 관련 연구

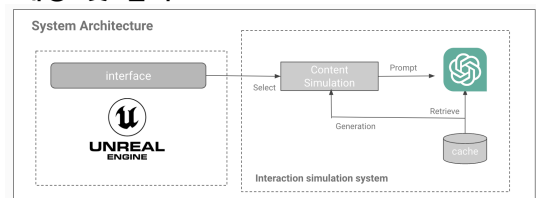
게임 시스템에서 LLM의 활용은 다양한 응용과 발전을 보여주고 있다. 게임 분야 있어, LLM은 NPC와의 상호작용,

전략적 게임 플레이, 게임 파이프라인 개선, 개인화된 게임 경험 등에서 적용되고 있다.[3][4][5]

NPC와의 상호작용에 있어, LLM은 프롬프트 엔지니어링을 통해 롤플레이팅 비디오 게임에 있어 플레이어의 몰입도를 높이는 효과를 보여주었다.[1] LLM을 활용한 NPC Dialogue System에서는 Scene-specific importation을 통해 다중 NPC에게 각각의 Personality, World info, Occupation 등을 할당하였다. 플레이어는 음성을 통해 프롬프트를 생성하고, 퀘스트를 위한 독백을 생성한다.

전략적 게임 플레이 사례에서도 LLM이 활용되고 있다. 관련 논문에서는 LLM을 활용하여, 기존 MAB 전략에 통합하여 전체적인 효율성을 높이는 'LLM-정보 전략'을 제안한다. 이는 복잡한 시나리오를 이해하고 정보에 기반한 결정을 내릴 수 있는 LLM의 능력을 활용하여 보상 분포에 적절하게 적응하지 못하는 문제를 해결하는 방법을 제안하였다[2].

### 3. 연구 내용 및 결과



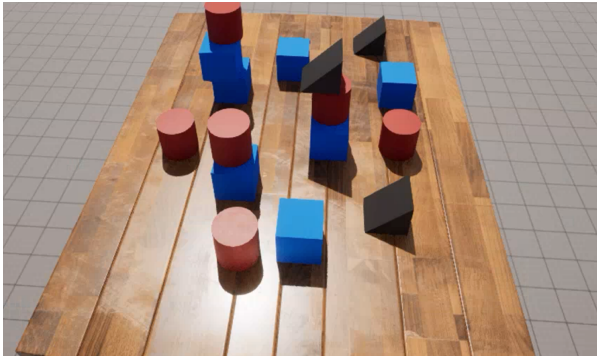


그림 1. 언리얼 엔진기반 시스템 구성도 및 스크린샷 이미지

위 이미지는 LLM을 활용하여 게임 시나리오에 맞는 디스크립션을 생성하는 기법을 적용한 게임 시뮬레이션 시스템을 제시하고 있으며, 전체적인 시스템은 (1) 콘텐츠를 선택하는 Select와 (2) 프롬프트 입력, (3) 기존 히스토리를 기반으로 한 Retrieval 및 응답 생성, (4) 게임엔진 통합으로 구성하였다.

첫 번째 단계에서는, 게임 콘텐츠를 선택 과정에서 선택한 게임의 정보들을 불러온다. 두 번째, 해당 콘텐츠 내용을 기반으로 사용자가 게임 진행방식에 대해 프롬프트를 입력하게 된다. 세 번째로는, 이전 대화 이력을 통해, 게임의 진행 사항과 접근 방식을 검색하고, 프롬프트에 대한 응답을 생성하게 된다. 네 번째는, 생성된 응답이 언리얼 엔진과 연동되어 발화를 통해 게임에서 출력되는 과정으로 구성된다.

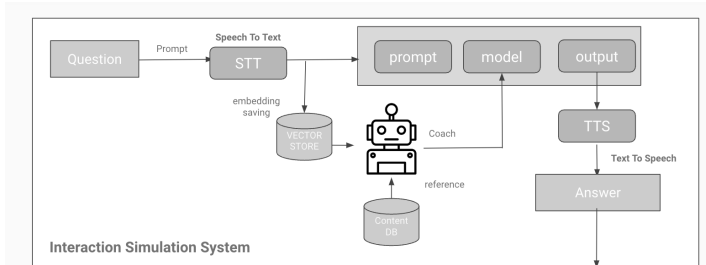


그림 2. LangChain기반 RAG을 사용한 음성 발화 자동성 구조

위 이미지는 게임 시나리오를 기반으로 STT(Speech To Text)를 사용하여 음성 정보로부터 프롬프트를 생성하고, TTS(Text To Speech)를 통해 음성 발화를 얻는 구조를 보여준다.

위 구조에서는 제안하는 기법은 RAG을 활용하여, 이전 프롬프트를 기반으로 게임 콘텐츠와 관련한 참조 문장에 대한 검색을 수행하며, GPT4-turbo 모델을 통해 프롬프트에 대한 디스크립션을 생성하게 된다. 여기서, 에이전트는 이전 프롬프트를 기반으로 게임에 대한 맥락 추론을 통해 게임 진행에 필요한 정보를 제공하는 역할을 수행한다.

4. 실험 결과 분석

명확한 결과 분석과 성능 평가를 위해, 게임 콘텐츠 내용을 포함하고 있는 참조 문장(Reference)과, 프롬프트로 사용되는 질의 문장(Question), 참조 문장을 사용하여 만들어진 생성 문장(Answer)로 나누어 구성하였으며, 문맥 추론과 문장 이해력을 테스트 하기 위한 집중력 향상 게임이라는 주제에 따라, (a) 도형 추론, (b) 수학 추론, (c) 상식 추론으로 나누고, 생성 모델을 이용하여 각 영역에서 10개 세트의 참조 문장과 생성 문장으로 구성하였다.

표1. 테스트셋 구성 예시

참조 문장 (Reference)	“다음 테이블 앞의 놓여진 도형들에 대해서, 반으로 접었을 때 완벽하게 하나가 되는 도형을 찾으세요.”
질의 문장 (Question)	“도형을 반으로 접었을 때, 같은 도형에는 뭐가 있지?”
생성 문장 (Answer)	“어떤 도형이 한 축을 중심으로 접었을 때, 양쪽 모양이 정확히 일치하면 그 도형은 축 대칭이라고 할 수 있어요. 한번 생각해 볼까요, 종이를 접었을 때 완벽하게 일치하는 도형은 무엇이 있을까요?”

표2. 프롬프트와 생성문장 예시

구분	Question	Answer
도형 추론 (a)	“도형을 반으로 접었을 때 같은 도형에는 뭐가 있지?”	어떤 도형이 한 축을 중심으로 접었을 때, 양쪽 모양이 정확히 일치하면 그 도형은 축 대칭이라고 할 수 있어요. 한 번 생각해 볼까요, 종이를 접었을 때 완벽하게 일치하는 도형은 무엇이 있을까요?
수학 추론 (b)	“루트의 여섯제곱을 계산하려면 어떻게 해야하지?”	“먼저 숫자의 제곱근을 구하고, 제곱근의 제제곱을 구하여 답을 얻을 수 있습니다.”
상식 추론 (c)	“사각형의 바로 왼쪽에 있는 도형은 무엇이지?”	“삼각형은 사각형의 오른쪽에 있고, 원통모양은 삼각형의 바로 왼쪽에 있다고 했습니다. 따라서 사각형의 바로 왼쪽에 있는 도형은 삼각형입니다.”

생성 문장의 정확한 디스크립션 생성을 위해 프롬프트 템플릿을 기초하였으며, 프롬프트 템플릿의 주요 구성은 에이전트의 역할과 태스크, 형식 및 제약 지정으로 구성하였다. 프롬프트는 퀴즈의 정답을 찾기 위한 질의(Question)

를 입력으로 사용하고, 생성된 디스크립션을 통해 정답을 찾기 위한 디스크립션을(Answer) 제공하며, 해당 퀴즈의 질문을 참조하여(Reference) 템플릿의 형식과 제약 지정에 따라 실행된 것을 보여준다.

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-\frac{r}{c})} & \text{if } c \leq r \end{cases}$$

$$BLEU = BP \times \exp \left( \sum_{n=1}^N w_n \log p_n \right)$$

$$R_{skip2} = \frac{SKIP2(X, Y)}{C(m, 2)}$$

$$P_{skip2} = \frac{SKIP2(X, Y)}{C(n, 2)}$$

$$F_{skip2} = \frac{(1 + \beta^2) R_{skip2} P_{skip2}}{R_{skip2} + \beta^2 P_{skip2}}$$

그림.5. BLEU(Bilingual Evaluation Understudy Score)와 ROUGE(Recall-Oriented Understudy for Gisting Evaluation)

위에서 구성한 데이터셋에 따라, LLM의 문맥 추론과 참조 문장과 생성문장의 유사성을 평가하기 위해, BLEU Score와 ROUGE-S Score를 사용하였다.[8][9] BLEU는 모델이 생성한 텍스트와 참조 텍스트 간의 n-gram 겹침을 측정하여 유사도를 평가하는 지표로 두 문장 간의 유사성을 측정한다. 또한, ROUGE-S은 Skip-Bigram을 기반으로 한 지표로, 두 단어가 텍스트 내에서 순서를 유지하면서 일정 간격으로 등장하는지 측정한다.

	BLEU	ROUGE-S
(a)	0.42	0.38
(b)	0.32	0.27
(c)	0.46	0.41

각 영역의 차이점 비교를 위해, 지표는 각 쌍들에 대한 평균값을 사용하였으며, 참조 문장과 디스크립션 생성 품질을 평가를 위해, 사용한 BLEU score는 (a)영역인 도형 추론에서 0.41, (b)영역인 수리 추론에서 0.32, (c)영역인 상식 추론에서 0.46의 유사성을 보였다. 또한, ROUGE-S Score의 경우, (a)영역인 도형 추론에서 0.38, (b)영역인 수리 추론에서 0.27, (c)영역인 상식 추론에서 0.41의 유사성을 보였다.

#### 4. 결론 및 향후 연구

게임 환경에 있어, 게임 플레이의 중단이나 인터페이스

의 방해는 사용자의 몰입감을 떨어트리는 문제가 될 수 있다. 튜토리얼 기반 게임 플레이는 사용자가 겪을 수 있는 게임의 기본 메커니즘이나 실마리에 대한 디스크립션을 통해 플레이어가 겪는 혼란을 감소시킬 수 있다.

이에, 본 논문에서는 실시간 게임 시뮬레이션 시스템에서 LLM을 활용한 게임 디스크립션 생성 기법을 제안한다. 게임 디스크립션 생성 기법은 게임 시나리오를 바탕으로 프롬프트를 생성하고, 이를 LLM에 질의하여, 게임 진행에 필요한 가이드라인을 생성한다. 제안하는 방식의 RAG 구조는 이전 프롬프트 내용을 기반으로, 문맥을 추론하고 텍스트를 생성하게 함으로써, 명확한 디스크립션을 제공하게 된다.

제안한 기법의 성능을 평가하기 위해, 세가지 영역에서 10 세트의 생성된 디스크립션과 참조 문장을 비교한 결과 약 0.4의 유사성을 보였다. 다만, 면밀한 문맥 추론 성능 평가와 특정 사례에서 reference의 정답이 직접 언급되거나, 논리적이지 않는 추론을 제공하는 경우에 대한 앞으로 추가적인 실험과 튜닝이 필요할 것으로 보인다.

#### 사사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송혁신인재양성(메타버스융합대학원)사업연구결과과로 수행되었음(IITP-2024-RS-2023-00256615)

#### 참고 문헌

- [1] Csepregi, Lajos Matyas. "The effect of context-aware llm-based npc conversations on player engagement in role-playing video games." *Unpublished manuscript*, 2021.
- [2] de Curtò, J., et al. "LLM-Informed Multi-Armed Bandit Strategies for Non-Stationary Environments." *Electronics* 12.13, 2023.
- [3] Shao, Yunfan, et al. "Character-llm: A trainable agent for role-playing." *arXiv preprint arXiv:2310.10158*, 2023.
- [4] Gallotta, Roberto, et al. "Large Language Models and Games: A Survey and Roadmap." *arXiv preprint arXiv:2402.18659*, 2024.
- [5] Wang, Zekun Moore, et al. "Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models." *arXiv preprint arXiv:2310.00746*, 2023.
- [6] Yao, Shunyu, et al. "React: Synergizing reasoning and acting in language models." *arXiv preprint arXiv:2210.03629*, 2022.
- [7] Schmidt, Douglas C., et al. "Towards a catalog of prompt patterns to enhance the discipline of prompt engineering.", 2023.
- [8] Papineni, Kishore, et al. "BLEU: a method for automatic evaluation of machine translation." *Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics*, 2002.
- [9] Lin, Chin-Yew. "ROUGE: A Package for Automatic Evaluation of Summaries." *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop. Vol. 8*. 2004.