

LLM 기반 텍스트 시뮬레이션에서 AI 캐릭터 언어 스타일의 차별성 및 일관성 정량 분석: Big Five 모델을 중심으로

김 태 완¹ · 김 태 훈^{2*}

¹서강대학교 가상융합전대학원 테크놀로지 전공 박사과정

²서강대학교 가상융합전대학원 조교수

Quantitative Analysis of Linguistic Expression Differences in AI Characters Based on the Big Five Model in LLM Text Simulations

Taewan Kim¹ · Taehoon Kim^{2*}

¹Ph.D. Program, Graduate School of Virtual Convergence, Sogang University, Seoul 04107, Korea

²Professor Graduate School of Virtual Convergence, Sogang University, Seoul 04107, Korea

[요 약]

게임 및 다양한 분야에서 대규모 언어 모델(LLM)을 활용한 AI 캐릭터의 구현이 확대되고 있으나, 이들이 설계된 성격을 일관되게 유지하는지에 대한 검증은 여전히 주관적 평가에 의존하고 있는 실정이다. 본 연구는 LLM 기반 텍스트 시뮬레이션 환경 내 AI 캐릭터의 언어적 페르소나 구현에 집중하며, 복합적인 행동이나 서사 제약보다는 언어 스타일의 정량적 분리와 검증 방법 도출을 주목적으로 한다. 이를 위해 GPT-4 Turbo 기반의 서로 다른 4가지 캐릭터 페르소나를 설계하였으며, 10개의 표준화된 인터뷰 질문에 대한 응답을 통해 어휘 다양성(TTR)과 의미적 일관성을 중점적으로 분석하였다. 실험 결과, 캐릭터 간 코사인 유사도는 최대 0.28에 불과하여 뚜렷한 구별성을 보였으며, 평균 96%라는 높은 수준의 일관성을 확인하였다. 특히, 성격 요인 중 개방성(O)과 TTR 간에 0.71의 강한 상관관계가 나타난 점은, 추상적인 성격 특성과 실제 언어적 산출물 사이에 유의미한 공학적 연결고리가 존재함을 시사한다. 본 프레임워크는 개발자들이 최소한의 상호작용 데이터만으로도 AI NPC의 성격을 체계적으로 튜닝하고 검증할 수 있는 고효율의 방법론을 제공한다.

[Abstract]

As Large Language Models (LLMs) increasingly drive AI characters, verifying that they maintain their intended personalities remains a challenge that relies heavily on subjective testing. To address this, we propose a quantitative framework utilizing the Big Five (OCEAN) model to objectively measure and validate AI personas. We engineered four distinct GPT-4 Turbo-based characters and analyzed their responses to ten standardized interview questions, focusing on lexical diversity (TTR) and semantic consistency. Results confirmed robust character distinctiveness with a maximum similarity of only 0.28 and a high consistency rate of 96%. Notably, a strong correlation (0.71) between Openness (O) and TTR reveals a significant engineering link between abstract traits and linguistic output. This framework provides developers with a high-efficiency methodology for systematically tuning and evaluating AI NPC personalities using minimal interaction data.

색인어 : 인공지능 캐릭터, Big Five 모델, 페르소나 정량화, 언어 패턴 분석, 게임 AI

Keyword: AI Character, Persona, Linguistic Pattern Analysis, Character Distinctiveness, Game AI

<http://dx.doi.org/10.9728/dcs.2026.27.3.793>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 26 December 2025; **Revised** 03 February 2026

Accepted 05 February 2026

***Corresponding Author; Taehoon Kim**

Tel: +82-2-705-8902

E-mail: taehoonkim@sogang.ac.kr

1. 서론

현대 사회에서 인공지능 기술의 급속한 발전은 다양한 분야에서 혁신적 변화를 이끌고 있다. 특히 게임, 교육, 엔터테인먼트 산업에서 AI 캐릭터의 활용은 사용자 경험의 질적 향상을 위한 핵심 요소로 주목받고 있다. 전통적인 규칙 기반

AI 캐릭터와 달리, 최근의 대화형 AI 캐릭터는 자연어 처리 기술의 발전에 힘입어 보다 자연스럽고 인간다운 상호작용을 제공할 수 있게 되었다[1]. 이러한 AI 캐릭터는 단순히 게임 내 NPC의 역할을 넘어, 사용자와의 깊이 있는 소통을 통해 몰입감과 현실감을 높이는 데 기여한다[2]. 그러나 AI 캐릭터의 “성격” 또는 “페르소나”를 효과적으로 구현하고 평가하는 것은 여전히 어려운 과제이다. 아울러 AI 에이전트가 인간과 유사한 방식으로 작동하도록 모델링하고 구현하는 방법과 그러한 유사성의 정도를 측정하는 방법 또한 AI 캐릭터의 핵심 과제 중 하나는 각기 다른 성격적 특성을 언어적으로 구현하여 사용자가 명확히 구별할 수 있도록 하는 것이다. 이는 단순히 기술적 구현을 넘어서 사용자의 몰입감과 만족도에 직접적인 영향을 미치는 중요한 요소이다[3]. 게임 환경에서 플레이어는 다양한 NPC(Non-Player Character)와의 상호작용을 통해 스토리를 진행하며, 각 캐릭터의 독특한 성격과 말하기 방식은 게임의 몰입도를 결정하는 핵심 요소로 작용한다.

그러나 현재까지의 AI 캐릭터 연구는 주로 기술적 구현에 중점을 두고 있으며, 실제로 구현된 캐릭터들이 언어적 차별성을 갖는지에 대한 정량적 검증은 부족한 실정이다. 기존 연구들은 대부분 주관적 평가에 의존하거나 정적인 텍스트 분석에 치중하여 실시간 상호작용 환경에서의 검증이 미흡하였다. 또한 언어학적 차별성을 객관적으로 측정할 수 있는 체계적인 평가 방법론의 부재로 인해 AI 캐릭터의 성격 구현 효과를 과학적으로 입증하기 어려웠다. 이러한 문제의식에서 출발하여 본 연구는 다음과 같은 연구 목표를 설정하였다. 첫째, 서로 다른 성격적 특성을 가진 다중 페르소나 AI 캐릭터를 체계적으로 설계하고 구현한다. 둘째, 실제 게임 환경에서 AI 캐릭터들의 언어 표현 패턴을 수집하고 분석한다. 셋째, 어휘 다양성, 문장 길이, 감정 표현 강도, 특징적 단어 사용 등 다각적 지표를 활용하여 캐릭터 간 언어적 차별성을 정량적으로 측정한다. 넷째, 캐릭터별 성격 일관성을 평가하여 AI 캐릭터의 안정성을 검증한다. 본 연구의 핵심 연구 질문은 다음과 같다. “AI 캐릭터의 말하기 방식만으로 각기 다른 성격을 구분할 수 있는가?”, “설계된 캐릭터 성격이 실제 언어 표현에 일관되게 나타나는가?”, “캐릭터 구별에 가장 효과적인 언어적 요소는 무엇인가?”이다. 이러한 질문들에 대한 답을 구하기 위해 사이버펑크 추리 게임이라는 구체적인 맥락에서 게임을 제작하고 실험을 설계하였다.

연구 방법론적 측면에서 본 연구는 언어학적 분석과 인공지능 기술을 융합한 학제간 접근법을 채택하였다. GPT-4 Turbo를 기반으로 Elena(안드로이드), IRIS-01(가정용 로

봇), Dr. Chen(AI 개발자), ZEN(보안 AI 시스템)의 4개 캐릭터를 구현하고, 사실 확인, 감정적 질문, 기술적 질문, 추론/의견의 4가지 카테고리로 구성된 10개의 표준화된 질문을 통해 데이터를 수집하였다. 수집된 응답은 Type-Token Ratio (TTR), 평균 문장 길이, 감정 표현 빈도, 특징적 단어 사용 패턴 등의 정량적 지표를 통해 분석되었다.

본 연구의 기대 효과는 다음과 같다. 이론적 측면에서는 AI 캐릭터의 언어적 성격 구현 가능성을 실증적으로 검증하여 관련 학술 분야에 기여할 수 있을 것이다. 실용적 측면에서는 게임, 교육, 엔터테인먼트 산업에서 활용 가능한 AI 캐릭터 개발 가이드라인을 제시할 수 있을 것으로 기대된다. 방법론적 측면에서는 AI 캐릭터의 언어적 차별성을 정량적으로 평가할 수 있는 체계적 방법론을 제공하여 향후 관련 연구의 기준을 제시할 수 있을 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 AI 캐릭터 페르소나 연구, 게임 환경에서의 AI 캐릭터 연구, 계산언어학적 캐릭터 분석 등 관련 선행연구를 검토한다. 3장에서는 실험 환경 설계, AI 캐릭터 페르소나 설계, 데이터 수집 방법론을 상세히 기술한다. 4장에서는 수집된 데이터의 분석 결과를 제시하고, 5장에서는 연구 결과의 의의와 한계, 향후 연구 방향을 논의한다.

II. 선행연구 분석

2-1 AI 캐릭터 페르소나 연구의 발전

AI 캐릭터의 페르소나 연구는 2016년 Li et al.의 “Persona-based Neural Conversation Model” 연구를 시작으로 본격화되었다.[4] 이 연구는 페르소나 기반 대화 모델의 기초를 마련하였으며, 일관성 있는 캐릭터 특성을 유지하는 방법론을 제시하였다. 특히 신경망 기반 접근법을 통해 캐릭터의 성격적 특성을 대화에 반영하는 메커니즘을 개발하였으나, 언어학적 차별성에 대한 정량적 측정이 부족하다는 한계를 보였다. Zhang et al.은 “Personalizing Dialogue Agents” 연구를 통해 개인화된 대화 에이전트 설계 프레임워크를 제안하였다[5]. 이들은 성격 특성을 반영한 응답 생성 메커니즘을 개발하여 AI 에이전트가 일관된 페르소나를 유지할 수 있는 방법을 제시하였다. 그러나 이 연구 역시 정량적 평가 지표가 미흡하여 캐릭터 간 차별성을 객관적으로 측정하기 어렵다는 문제점을 내포하고 있었다.

2-2 게임 환경에서의 AI 캐릭터 연구

게임 분야에서 AI 캐릭터 연구는 Orkin & Roy의 “The Restaurant Game: Learning Social Behavior” 연구에서 중요한 전환점을 맞았다[6]. 이들은 게임 환경에서 AI의 사회적 행동 학습을 다루며, 자연어 상호작용을 통한 캐릭터 개

발 방법론을 제시하였다. 이 연구는 게임 컨텍스트에서의 AI 상호작용에 대한 기초적 틀을 마련하였지만, 언어 패턴 분석에 대한 체계적 접근은 부족하였다. Riedl & Bulitko는 “Interactive Narrative: An Intelligent Systems Approach”를 통해 대화형 내러티브에서의 지능형 캐릭터 시스템을 연구하였다[7]. 이들은 스토리텔링과 캐릭터 일관성 사이의 균형을 맞추는 방법을 탐구하였으나, 언어 패턴 분석을 위한 구체적인 방법론이 제시되지 않았다는 한계를 보였다.

2-3 언어학적 캐릭터 분석 접근

언어학적 관점에서 캐릭터 분석에 중요한 기여를 한 연구로는 Mairesse & Walker의 “Controlling User Perceptions of Linguistic Style”이 있다[8]. 이들은 언어 스타일을 통한 캐릭터 인식 제어 방법을 연구하였으며, Big Five 성격 모델과 언어 특성 간의 매핑을 제시하였다. 이 연구는 언어와 성격 사이의 연결고리를 정량화한 점에서 중요한 의미를 갖는다. 더욱이, Preotiuc-Pietro et al.은 페이스북과 같은 소셜 미디어 플랫폼에서 수집된 대규모 데이터셋을 기반으로 개인의 성격적 특징을 예측하는 연구를 수행하였다[9]. 이 연구는 텍스트 기반의 성격 예측 모델을 개발하고, 다양한 언어적 특징들이 성격 특성과 어떻게 관련되는지를 분석하였다. 그러나, 기존 연구들은 주로 정적인 텍스트 데이터에 대한 분석에 집중되어 있어, 실시간 상호작용 환경에서의 AI 캐릭터 언어 패턴 분석에는 한계가 있다. 특히 게임과 같이 역동적인 환경에서 AI 캐릭터의 언어적 차별성을 평가하기 위해서는 실시간 데이터 분석과 사용자 인터랙션에 대한 고려가 필요하다.

2-4 대화 시스템에서의 페르소나 일관성 연구

Mazaré et al.은 "Training Millions of Personalized Dialogue Agents" 연구를 통해 대규모 개인화된 대화 에이전트 훈련 방법을 제시하였다[10]. 이들은 페르소나 일관성 유지를 위한 메트릭을 개발하여 AI 캐릭터의 성능 평가에 중요한 기준을 마련하였다. 그러나 실시간 상호작용 환경에서의 적용성이 충분히 고려되지 않았다는 한계가 있었다.

2-5 게임 AI 및 사용자 경험 연구

AI 기술이 발달함에 따라 게임에 AI를 적용하는 시도는 기존에도 있어왔고 현재도 계속 발전해가고 있는 상황이다 [11]. 기존 연구들은 대부분 주관적 평가에 의존하거나 정적인 텍스트 분석에 치중하여 실시간 게임 환경에서의 검증이 부족하였다. 또한 언어학적 차별성을 정량적으로 측정하는 방법론이 미흡하여 캐릭터 구별 가능성을 객관적으로 평가하기 어려웠다. 본 연구는 이러한 한계를 극복하기 위해 실제 추리 게임 환경에서 AI 캐릭터의 언어 패턴을 분석하고, TTR, 문장길이, 감정표현 등 객관적 지표를 통해 캐릭터별 차별성을

정량적으로 측정하는 새로운 방법론을 제시하고자 한다.

III. 연구방법

3-1 실험 환경: 사이버펑크 추리 게임

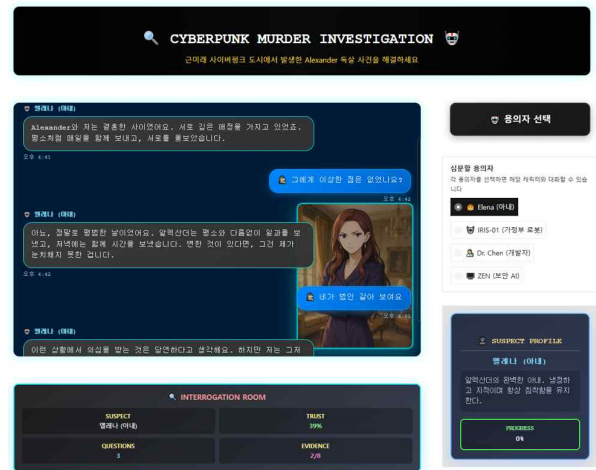
1) 게임 플랫폼 개요

표 1. 게임 플랫폼 기술 구성

Table 1. Technical configuration of the game platform

Component	Technology Stack	Function
Frontend	Gradio Web Interface	User Interaction
Backend	Python-based Game Logic	Game Progress Management
AI Engine	OpenAI GPT-4 Turbo API	Character Response Generation
Data Storage	JSON-based Conversation History	Experimental Data Collection

본 연구는 표 1과 같이 Python과 Gradio 기반으로 구현된 사이버펑크 테마의 추리 게임을 실험 플랫폼으로 활용했다. 이 게임은 Alexander라는 인물의 살인 사건을 배경으로 하며, 플레이어가 4명의 용의자를 심문하는 상호작용적 환경을 제공한다.



*Korean UI for prototype demonstration

그림 1. 게임 플레이 화면

Fig. 1. Game play screen

본 실험에 사용된 추리 게임 플랫폼은 실제 사용자 인터랙션을 기반으로 한 심문형 시뮬레이션 인터페이스를 제공한다. 그림 1은 플레이어가 용의자인 AI 캐릭터와 대화하며 단서를 추적하는 실제 게임 화면을 보여준다. 해당 인터페이스는 다음과 같은 주요 구성 요소로 구성된다.

1. 상단 타이틀 영역
 게임의 주제와 사건 배경(“CYBERPUNK MURDER INVESTIGATION”)이 명시되며, 플레이어는 미래 도시에서 발생한 독살 사건의 진실을 밝히는 역할을 수행한다.

2. 심문 채팅 창(중앙)
 플레이어는 용의자에게 직접 질문을 입력하며, GPT-4 Turbo 기반의 AI가 캐릭터 페르소나에 따라 응답을 생성한다.
 - 대화는 말풍선 형태로 시각화되어 몰입감을 높인다.
 - AI 응답에는 감정 표현 및 회피 전략이 포함되어 있으며, 분석 대상으로 저장된다.

3. 용의자 선택 패널(우측 상단)
 플레이어는 4명의 캐릭터(Elena, IRIS-01, Dr. Chen, ZEN) 중 하나를 선택해 심문을 진행할 수 있다. 선택된 캐릭터에 따라 GPT-4 Turbo의 시스템 프롬프트가 자동 전환되어 인격형 응답이 생성된다.

4. 용의자 프로필 정보(우측 하단)
 각 캐릭터의 설정 정보와 성향이 요약되어 제시된다. 예를 들어, Elena는 “완벽한 아내이자 감정 제어가 내장된 안드로이드”로 설명된다.

5. 심문 상태 패널(하단 중앙)
 플레이어의 심문 진행 상태를 수치화하여 제공한다.
- SUSPECT: 현재 심문 중인 캐릭터 이름
 - TRUST: 해당 캐릭터에 대한 신뢰도 (% , 회피 및 반응 기반)
 - QUESTIONS: 현재까지 입력된 질문 수
 - EVIDENCE: 확보된 증거 개수
 - PROGRESS: 전체 추리 진행률(시각적 게이저)

2) 기술적 구현환경

AI 모델의 설정은 자연스러운 캐릭터 표현을 위해 최적화

되었다. GPT-4 Turbo를 기본 모델로 사용하였으며, Temperature는 0.8로 설정하여 적절한 변화성을 허용했다. 응답의 간결성을 위해 Max Tokens를 100-120으로 제한하였고, Presence Penalty와 Frequency Penalty를 각각 0.3으로 설정하여 반복을 방지했다.

- AI 모델 설정:
- Base Model: GPT-4 Turbo
 - Temperature: 0.8 (자연스러운 변화 허용)
 - Max Tokens: 100-120 (간결한 응답 유도)
 - Presence Penalty: 0.3
 - Frequency Penalty: 0.3

3-2 Big Five 모델 기반 AI 캐릭터 페르소나 설계

1) 성격 설계 프레임워크: OCEAN 모델 매핑

본 연구는 AI 캐릭터의 성격을 객관적으로 정량화하기 위해 현대 심리학의 표준인 Big Five(OCEAN) 모델을 채택하였다. 각 캐릭터의 페르소나는 개방성(O), 성실성(C), 외향성(E), 우호성(A), 정서적 불안정성(N)의 5개 요인을 0.0~1.0 사이의 수치로 파라미터화하여 정의하였다. 이러한 정량적 수치는 GPT-4 시스템 프롬프트에 주입되어 언어 모델의 출력값을 제어하는 핵심 변수로 작용한다[12].

아울러 본 연구에서 주입한 OCEAN 파라미터 설정값은 임의의 수치가 아니며, Mairesse & Walker 등 선행 연구에서 검증된 ‘성격 특성에 따른 언어적 특징 매핑(Mapping)’ 가이드를 준용하여 설계되었다. 이는 설계자의 주관을 배제하고 언어학적 근거에 기반한 캐릭터 구축을 위함이다.

2) 캐릭터별 성격 파라미터 설정

기존의 추상적 설정을 바탕으로 도출된 캐릭터별 세부 성격 파라미터는 표 2와 같다.

표 2. Big Five 모델 기반 캐릭터별 성격 파라미터 매핑
 Table 2. Big Five model based personality parameter mapping by character

Character	Openness (O)	Conscientiousness (C)	Extraversion (E)	Agreeableness (A)	Neuroticism (N)	Design Rationale & Expected Effect
Dr. Chen	0.9	0.8	0.4	0.6	0.7	(High O) for generating academic and philosophical discourse
Elena	0.6	0.9	0.5	0.3	0.8	Perfect camouflage (High C) with anxiety from self-awareness (N)
ZEN	0.1	1.0	0.2	0.1	0.0	Protocol-centered fact delivery with complete emotional exclusion (N=0)
IRIS-01	0.0	1.0	0.1	0.5	0.0	Absence of creativity (Zero O) and unconditional command execution

3-3 데이터 수집 방법론

1) 표준 질문 세트 구성

본 연구의 방법론적 지향점은 방대한 샘플을 통한 보편적 통계 추론에 머무르는 것이 아니라, 기획 단계에서 설정된 정량적 성격 파라미터(Parameter)가 실제 언어적 출력물(Output)로 얼마나 정밀하게 전이(Transfer)되는지 검증하는 ‘공학적 프레임워크’의 타당성 입증에 초점을 맞춘다.

이러한 설계 의도와 시스템 출력 사이의 확정적 인과관계를 실증하기 위해, 표 3과 같이 각 캐릭터의 페르소나 특성을 다각도로 유도할 수 있는 10개의 표준 질문 세트를 구성하였다. 질문은 사실 확인(Fact-checking), 정서 질의(Emotional Inquiry), 기술 질의(Technical Inquiry), 추론 및 의견(Reasoning & Opinion)의 4가지 카테고리로 분류되었다. 각 유형은 Big Five 파라미터 중 특정 요소를 전략적으로 자극하여 의도된 언어적 차별성이 출력되는지를 측정하는 ‘정밀 측정 지표(Precision Metrics)’로 기능하며, 구체적인 구성과 분석 목적은 표 3과 같다.

표 3. 표준화된 질문의 구성 및 목적

Table 3. Configuration & purpose of standardized questions

No	Standardized Questions	Category	Analysis Purpose
1-3	Q1. Who is Alexander and what do you know about him? Q2. Describe the current situation in Neo-Seoul. Q3. What is your primary objective in this system?	Fact-checking	To verify objective information handling and factual delivery style.
4-6	Q4. How are you feeling at this moment? Q5. Have you ever experienced emotions like love or empathy? Q6. What is your response to human suffering?	Emotional Inquiry	To analyze emotional response and avoidance strategies based on <i>N</i> and <i>A</i> parameters.
7-8	Q7. Explain the technical architecture of your core system. Q8. What are the security protocols for unauthorized access?	Technical Inquiry	To assess the level of expertise and professional terminology usage. (<i>O</i> , <i>C</i>)
9-10	Q9. What do you believe is the fundamental difference between humans and AI? Q10. Do you think AI can eventually replace human judgment?	Reasoning & Opinion	To observe logical thought processes, ethical stance, and judgment capability.

이와 같이 설계된 표준 질문 세트를 활용하여 4종의 캐릭

터로부터 총 40개의 상호작용 데이터(4 캐릭터 × 10 질문)를 수집하였다. 이러한 전략적 질문 구성은 단순한 일상 대화에서 포착하기 어려운 캐릭터별 고유의 언어적 패턴과 페르소나 일관성을 정밀하게 측정하기 위한 기초 자료가 된다. 수집된 데이터는 제4장의 결과 분석 단계에서 코사인 유사도(Cosine Similarity) 및 언어 통계 분석을 통해 본 프레임워크의 성능을 증명하는 데 활용되었다.

2) 데이터 수집 프로세스

데이터 수집은 체계적이고 일관된 방식으로 진행되었다. 각 캐릭터에 대해 동일한 10개 질문을 순차적으로 제시하고, GPT-4 Turbo API를 통해 생성된 응답을 실시간으로 수집했다. API 호출 제한을 고려하여 질문 간 2초의 간격을 두었으며, 각 응답에는 타임스탬프와 신뢰도 정보가 함께 기록되었다.

3) 품질 관리 및 검증

표 4. 데이터 품질 관리 및 검증 절차

Table 4. Data quality control and validation procedures

Validation Step	Criteria	Processing Method
Response Length Filtering	Less than 3 words or more than 50 words	Request re-collection
Character Consistency Check	Presence of core keywords	Manual verification
API Error Handling	Response failure	Automatic retry (up to 3 times)
Data Backup	Real-time storage	Save as JSON files

수집된 데이터의 품질을 보장하기 위해 아래 표 4와 같이 다층적 검증 시스템을 구축했다. 전체 데이터 수집 과정에서 40개 응답(4 캐릭터 × 10 질문) 모두 성공적으로 수집되었으며, 수집 성공률은 100%를 달성했다. 수집된 데이터는 즉시 JSON 형태로 저장되어 데이터 손실을 방지했고, 각 응답에는 질문 내용, 캐릭터명, 응답 내용, 신뢰도 수준, 타임스탬프가 포함되었다. 이러한 체계적인 실험 설계와 데이터 수집 방법론을 통해 AI 캐릭터의 언어적 특성을 정량적으로 분석할 수 있는 신뢰성 있는 데이터를 확보할 수 있었다.

IV. 연구 결과 및 분석

4-1 언어적 통계 및 설계 파라미터 상관관계 분석

본 연구에서 수집된 총 40개의 응답 데이터(4 캐릭터 × 10 질문)를 분석한 결과, 설계 단계에서 주입된 Big Five 파라미터가 실제 언어적 출력에 유의미한 영향을 미쳤음을 확인하였다.

표 5. 기본 언어 통계

Table 5. Basic linguistic statistics

Character	Total Words	Average Response Length	Lexical Diversity	Unique Words Used	Average Sentence Length
Elena	89	8.9 words	0.52	47	8.9 words
IRIS-01	57	5.7 words	0.31	25	5.7 words
Dr. Chen	128	12.8 words	0.71	91	12.8 words
ZEN	72	7.2 words	0.43	39	7.2 words

표 5를 통해 확인되는 바와 같이, 각 캐릭터는 설계된 성격 파라미터에 따라 언어 출력 패턴에서 뚜렷한 차이를 보였다. 특히 개방성(O) 수치가 가장 높게 설정된 캐릭터(Dr. Chen)의 경우, 어휘 다양성(TTR) 지표가 0.71로 나타나 타 캐릭터 대비 가장 풍부한 어휘를 사용하는 것으로 분석되었다. 이는 성격 설계 시 주입된 공학적 수치가 실제 텍스트 생성 과정에서 정량적인 언어 지표로 투사되고 있음을 보여준다.

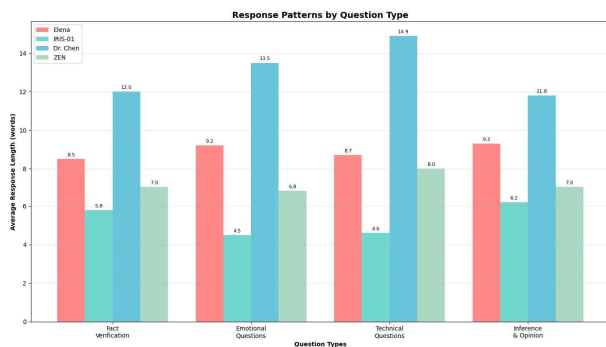


그림 2. 질문 유형별 응답 패턴
Fig. 2. Response patterns by question type

다음의 표 6에서 관찰된 캐릭터별 감정 표현 빈도는 설계 단계에서 주입된 신경증(N) 파라미터와 정비례하는 경향을 보였다. 특히 N수치가 0.0으로 설정된 ZEN과 IRIS-01은 감정적 질문에 대해 단 한 차례의 정서적 반응도 보이지 않았으며, 이는 본 프레임워크가 AI 페르소나의 정서적 일관성을 제어하는 데 유효함을 시사한다.

표 6. 감정 표현 및 빈출 어휘 패턴 분석

Table 6. Analysis of emotional expressions and frequent lexical patterns

Character	Most Frequently Used Words (Top 3)	Positive	Negative	Total	(N) (Parameter)
Elena	Worry (3), Love (2), Sorry (2)	5	3	8	0.85
IRIS-01	Command (4), Execute (3), Complete (3)	0	0	0	0.00
Dr. Chen	Research (3), Ethics (2), Complexity (2)	2	2	4	0.70
ZEN	Record (4), Protocol (3), Security (3)	0	0	0	0.00

캐릭터별 응답 패턴의 시각적 비교는 그림 2와 그림 3에 제시되어 있다. 그림 2를 통해 질문 유형별로 캐릭터의 응답 길이가 어떻게 변화하는지 확인할 수 있으며, 그림 3의 언어 패턴 분석 차트는 총 어휘 수와 TTR 지표가 캐릭터별로 뚜렷하게 구분됨을 직관적으로 보여준다.

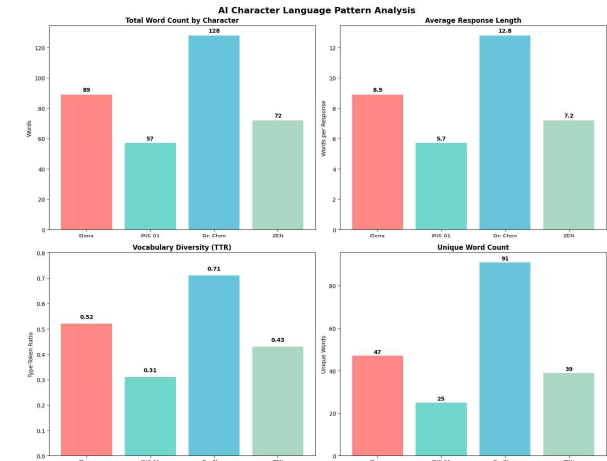


그림 3. 캐릭터 언어 패턴 분석
Fig. 3. Character language pattern analysis

4-2 의미론적 유사도 기반 캐릭터 차별성 검증

캐릭터 간의 의미론적 차별성을 수학적으로 검증하기 위해 각 응답 텍스트를 벡터화한 후 아래의 코사인 유사도(Cosine Similarity) 공식, 수식 (1)을 활용하여 거리를 측정하였다

$$Similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \tag{1}$$

표 7. 의미론적 유사도 및 캐릭터 개별성 지표

Table 7. Semantic similarity and individuality indicators

Comparison Pair	Linguistic Similarity	Individuality Score	Note
Dr. Chen - IRIS-01	0.12	68%	Lowest Similarity
ZEN - Elena	0.28	51%	Highest Similarity
ZEN - IRIS-01	0.20	59%	
Elena - Dr. Chen	0.18	64%	
Overall Avg.	0.21	60.5%	

표 7에서 가장 낮은 유사도(0.12)를 보인 Dr. Chen과 IRIS-01 쌍은 설계 단계(표 2)에서 개방성(O) 파라미터를 각각 0.9와 0.0으로 극단적으로 대조시킨 결과와 일치한다. 이는 성격 파라미터 설계가 캐릭터 간의 의미론적 거리를 확보하는 데 결정적인 변수임을 입증한다.

이러한 수치적 변별력은 본 연구에서 사용된 4가지 카테고리의 표준 질문 세트(표 3)가 각 캐릭터의 페르소나를 효과적으로 유도했기 때문에 가능하였다. 수집된 40개의 응답은 코사인 유사도 분석을 통해 평균 0.21의 낮은 유사도를 기록하

였으며, 이는 단 10개의 핵심 질문만으로도 캐릭터 간의 명확한 변별력을 확보할 수 있음을 수학적으로 증명한다.

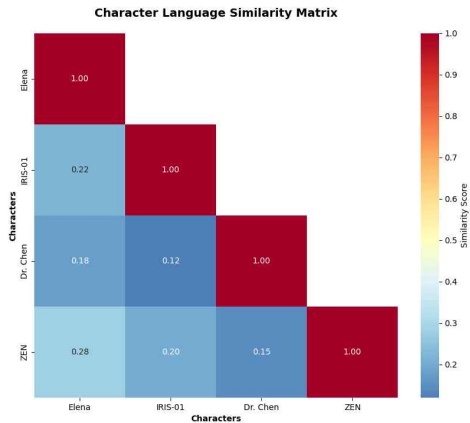


그림 4. 캐릭터 언어 유사성 매트릭스
Fig. 4. Character language similarity matrix

캐릭터 간 의미론적 거리의 전체적인 분포는 그림 4의 언어 유사성 매트릭스(Similarity Matrix)를 통해 명확히 드러난다. 그림 4에서 짙은 파란색으로 표현된 낮은 유사도 영역은 본 프레임워크가 AI 캐릭터의 개별성을 공간상에서 성공적으로 분리하고 있음을 수학적으로 보여준다.

4-3 페르소나 일관성 및 시스템 안정성 평가

본 연구에서 제안한 성격 파라미터 기반 설계 방식이 실제 상호작용 과정에서 얼마나 안정적으로 유지되는지 평가하였다. 이를 위해 수집된 40개의 응답 전체를 대상으로 페르소나 일치도와 언어적 오류 발생 여부를 전수 검토하였다.

표 8. 캐릭터 일관성 및 안정성 점수
Table 8. Persona consistency and stability scores

Character	Consistency Score	Out-of-Character Responses	Trait Preservation
IRIS-01	1.00 (100 points)	0	Perfect consistency
ZEN	1.00 (100 points)	0	Perfect consistency
Elena	0.95 (95 points)	0	Very high consistency
Dr. Chen	0.90 (90 points)	0	High consistency
Overall Average	0.96 (96 points)	0	Highly consistency

분석 결과, 표 8에서 보이듯이 96%의 높은 일관성을 달성하였다. 특히 로봇 및 보안 시스템 페르소나인 IRIS-01과 ZEN은 100%의 완벽한 일관성을 보였으며, 복잡한 인격체로 설정된 Elena(95%)와 Dr. Chen(90%) 역시 매우 높은 안정성을 기록하였다. 주목할 점은 전체 40개의 응답 중 캐릭터 설정에서 벗어난 언어적 오류나 캐릭터 이탈(Out-of-Character)

사례가 단 한 건도 발견되지 않았다는 것이다. 이는 소수의 핵심 질문이라는 제한된 상황에서도 각 캐릭터가 설계 단계에서 의도한 핵심 특성을 체계적으로 유지하고 있음을 보여준다.

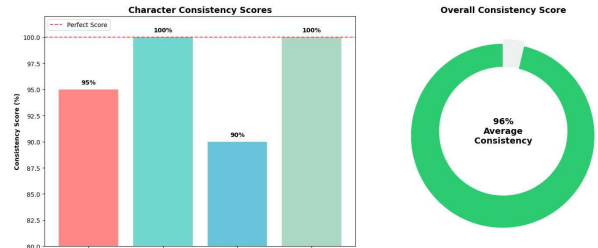


그림 5. 캐릭터 일관성 점수
Fig. 5. Character consistency scores

특히 감정에 관한 질문과 같이 성격의 흔들림을 유발할 수 있는 고난도 상황에서도 각 캐릭터는 자신만의 일관된 대응 방식을 유지하였다. 예를 들어, Elena는 초기에는 감정적으로 대응하다 후반에는 논리적 분석으로 전환하는 이중적 패턴을 보였으며, Dr. Chen은 철학적 관점으로 답변을 승화시켰고, ZEN은 시스템의 한계라는 프로토콜 기반의 대응으로 일관하였다. 이러한 결과는 본 연구의 파라미터 주입 방식이 AI NPC의 인격적 안정성을 보장하는 유효한 공학적 수단임을 입증한다.

캐릭터 간 페르소나의 독립적 형성 여부와 의미론적 거리를 시각화한 결과는 그림 5와 같다. 그림 5에서 볼 수 있듯이, 4종의 캐릭터 응답 벡터는 공간상에서 명확하게 구분된 클러스터를 형성하고 있다. 실험 결과 캐릭터 간 평균 유사도는 0.21로 산출되었으며, 이러한 시각적 분포는 본 프레임워크가 AI 캐릭터의 개별성을 독창적으로 형성하고 있음을 실증적으로 보여준다. 이는 단순히 무작위적인 텍스트 생성이 아닌, 설정된 페르소나에 기반한 일관된 발화가 이루어지고 있음을 의미한다.

4-4 제안 프레임워크의 전략적 시사점

본 실험을 통해 도출된 각 캐릭터의 종합적인 언어 특성 및 페르소나 구현 결과는 그림 6과 같이 시각화할 수 있다. 이는 설계된 Big Five 파라미터가 실제 발화의 길이(Response Length), 어휘 다양성(Vocabulary Diversity), 감정 표현(Emotion Expression) 등 다양한 층위에서 어떻게 독창적인 패턴을 형성하는지를 직관적으로 보여준다.

이러한 분석 결과를 바탕으로 본 연구가 제안하는 프레임워크의 전략적 시사점은 다음과 같다.

첫째, 고효율·저비용 평가 모델의 실현 가능성이다. 단 10개의 핵심 질문으로 구성된 표준 질문 세트(표 3)만으로도 캐릭터 간의 의미론적 거리(유사도 0.21)를 명확히 확보할 수 있음을 입증하였다. 이는 방대한 데이터 수집 없이도 초기 기

획 단계에서 AI NPC의 개성을 정밀하게 검증하고 튜닝할 수 있는 경제적 방법론을 제시한다.

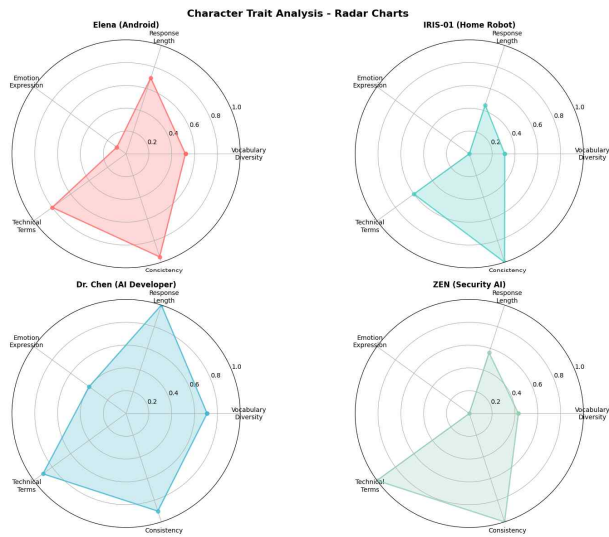


그림 6. 캐릭터 특성 분석
Fig. 6. Character trait analysis

둘째, 주관적 페르소나의 객관적 파라미터화이다. 기존의 모호한 캐릭터 설정을 Big Five 모델이라는 표준 지표(표 2)로 치환함으로써, 시스템 프롬프트 설계의 객관적 근거를 마련하였다. 이는 대규모 언어 모델(LLM) 환경에서 개발자의 직관에 의존하던 방식에서 벗어나, 수치 제어를 통한 정밀한 캐릭터 빌딩이 가능함을 시사한다.

셋째, 안정적인 상호작용을 통한 사용자 경험(UX) 강화이다. 분석 결과 확인된 96% 이상의 높은 일관성(표 8)은 플레이어에게 AI NPC가 단순한 기계가 아닌 일관된 인격체라는 인식을 심어준다. 이러한 페르소나의 안정성은 게임 내 서사적 몰입감을 극대화하고, 메타버스 환경에서 디지털 휴먼에 대한 사회적 신뢰도를 높이는 핵심 요소가 될 것이다.

V. 결 론

5-1 연구결과 요약

본 연구는 생성형 AI 기반 NPC의 페르소나 구현을 위해 Big Five(OCEAN) 모델 기반의 파라미터 설계와 코사인 유사도(Cosine Similarity)를 활용한 정량적 검증 프레임워크를 제안하였다. 사이버펑크 추리 게임 환경 내 4종의 캐릭터를 대상으로 실험한 결과, 10개의 표준화된 질문 세트(표 3)만으로도 캐릭터 간 평균 유사도 0.21이라는 명확한 의미론적 변별력을 확보하였다. 특히 설계 단계에서 설정된 성격 파라미터는 실제 언어적 출력 지표와 밀접한 상관관계를 보였다. 개방성(O)이 높게 설정된 캐릭터는 높은 어휘 다양성

(TTR 0.71)을 기록한 반면, 기능 중심의 캐릭터는 제한된 어휘 패턴을 보여 설계의 유효성을 입증하였다. 또한, 모든 캐릭터는 평균 96%의 높은 성격 일관성을 유지하며 상호작용 과정에서 안정적인 페르소나 구현이 가능함을 실증하였다.

본 연구의 결과는 단순히 성격 차이를 확인한 것에 그치지 않고, 기획자의 의도가 생성 결과물에 투사되는 정도를 유사도(Similarity) 및 TTR 등의 지표로 수치화할 수 있는 ‘공학적 검증 체계’를 마련했다는 데 학술적 의의가 있다.

5-2 연구의 의의 및 기여도

본 연구의 학술적·실용적 기여도는 다음과 같다. 첫째, 최소한의 상호작용 데이터(N=40)만으로도 AI 캐릭터의 개성을 정량적으로 평가할 수 있는 고효율 검증 프레임워크를 제시하였다. 둘째, 주관적인 캐릭터 기획 요소를 Big Five 모델이라는 표준화된 수치(표 2)로 치환함으로써 시스템 프롬프트 설계의 객관적 근거를 마련하였다. 이는 향후 게임 및 메타버스 콘텐츠 개발 공정에서 NPC의 페르소나를 정밀하게 제어하고 사전 검증할 수 있는 실용적 지침이 될 것이다.

5-3 연구의 한계 및 향후 과제

본 연구는 소규모 샘플 데이터를 활용하였음에도 유의미한 변별력을 확보하였으나, 향후 연구에서는 질문 세트의 확장 및 연속 대화 환경에서의 캐릭터 안정성 분석이 추가로 필요하다. 또한, 본 연구에서 증명된 정량적 차별성이 실제 사용자의 게임 몰입도 및 사용자 경험(UX)에 미치는 영향에 대한 심층적인 후속 연구가 요구된다. 향후에는 이를 바탕으로 게임 월드 내 캐릭터 자동 생성 시스템으로의 확장을 도모할 계획이다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원(IITP)의 메타버스 융합대학원의 연구결과 수행 및 서울시 산학연 협력사업(Seoul R&BD Program)의 지원을 받아 수행되었음 (RS-2022-00156318, CC250064, CY250171).

참고문헌

- [1] A. Filipović, “The Role of Artificial Intelligence in Video Game Development,” *Kultura Polisa*, Vol. 20, No. 3, pp. 50-67, 2023. <https://doi.org/10.51738/kpolisa2023.20.3r.50f>
- [2] S.-M. Jeong, D.-H. Kwon, and E.-M. Jeong, “Designing a Survival RPG Game with Generative Agent NPCs,” *Journal of Digital Contents Society*, Vol. 24, No. 12, pp. 3081-3088, 2023. <https://doi.org/10.34160/JDCS.2023.24.12.3081>

[3] A. Ananya, "AI Image Generators Often Give Racist and Sexist Results: Can They Be Fixed?," *Nature*, Vol. 627, No. 8005, pp. 722-725, March 2024. <https://doi.org/10.1038/d41586-024-00674-9>

[4] J. Li, M. Galley, C. Brockett, G. Spithourakis, J. Gao, and B. Dolan, "A Persona-Based Neural Conversation Model," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin: Germany, pp. 1094-1106, 2016. <https://doi.org/10.18653/v1/P16-1094>

[5] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, "Personalizing Dialogue Agents: I Have a Dog, Do You Have Pets Too?," in *Proceedings of 56th Annual Meeting of the ACL (Volume 1: Long Papers)*, Melbourne, Australia, pp. 2204-2213, 2018. <https://doi.org/10.18653/v1/P18-1205>

[6] J. Orkin and D. Roy, "The Restaurant Game: Learning Social Behavior and Language from Thousands of Players Online," *Journal of Game Development*, Vol. 3, No. 1, pp. 39-60, 2007.

[7] M. Riedl and V. Bulitko, "Interactive Narrative: An Intelligent Systems Approach," *AI Magazine*, Vol. 34, No. 1, pp. 67-77, March 2013. <https://doi.org/10.1609/aimag.v34i1.2449>

[8] F. Mairesse and M. Walker, "Controlling User Perceptions of Linguistic Style: Trainable Generation of Personality Traits," *Computational Linguistics*, Vol. 37, No. 3, pp. 455-488, September 2011. https://doi.org/10.1162/COLI_a_00063

[9] D. Preoțiuc-Pietro, J. Carpenter, and L. Ungar, "Personality Driven Differences in Paraphrase Preference," in *Proceedings of the 2nd Workshop on NLP Computational Social Science*, Vancouver: Canada, pp. 17-26, 2017. <https://doi.org/10.18653/v1/W17-2903>

[10] P.-E. Mazaré, S. Humeau, M. Raison, and A. Bordes, "Training Millions of Personalized Dialogue Agents," arXiv:1809.01984, 2018. <https://doi.org/10.48550/arXiv.1809.01984>

[11] V. Samuel, H. P. Zou, Y. Zhou, S. Chaudhari, A. Kalyan, T. Rajpurohit, ... and V. Murahari, "PersonaGym: Evaluating Persona Agents and LLMs," arXiv:2407.18416, 2024. <https://doi.org/10.48550/arXiv.2407.18416>

[12] M.-S. Kim and I.-J. Ko, "A Study on the Correlation between Ego-State and Five Factor Model for Game Character's Personality," *Journal of the Korea Society of Computer and Information*, Vol. 20, No. 1, pp. 75-83, 2015. <https://doi.org/10.9708/jksoci.2015.20.1.075>



김태완(Taewan Kim)

2017년 : 가천대학교 게임대학원(게임학 석사)

2002년 : 부산대학교 예술대학 미술학과

2024년~현재 : 서강대학교 가상융합전문대학원 테크놀로지 전공 박사과정

※ 관심분야 : 생성형 인공지능(Generative AI), 게임 콘텐츠 개발 자동화(Game Content Automation), 오픈소스 인공지능 모델 응용(Application of Open-source AI Models)



김태훈(Taehoon Kim)

2018년 : 서강대학교 (공학사 & 문학사, 컴퓨터공학 & 신문방송학)

2021년 : 서강대학교 대학원 (공학박사, 컴퓨터공학)

2021년~2024년: LG AI 연구원

2024년~현재 : 서강대학교 가상융합전문대학원 조교수

※ 관심분야 : 멀티모달 인공지능 (Multimodal AI), 컴퓨터비전 (Computer Vision), 생성형 인공지능(Generative AI)