일반논문 (Regular Paper)
방송공학회논문지 제31권 제4호, 2026년 7월 (JBE Vol.31, No.4, July 2026)
https://doi.org/10.5909/JBE.2026.31.4.722
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

# 상황인지 기반 사회적 촉매(Social Catalyst)형 AI 컴패니언 시스템 아키텍처 설계 및 최적 LLM 엔진 선정을 위한 비교 연구

강 자 원[a], 최 승 관[a]‡

# System Architecture Design of a Situation-Aware AI Companion as a Social Catalyst and a Comparative Study for Optimal LLM Selection

Jawon Kang[a] and Seungkwan Choi[a]‡

## 요 약

본 연구는 기술 발전이 인간 소외를 야기하는 '함께 있지만 따로 있는(Alone Together)' 역설에 주목하여, AI를 인간 간 상호작용을 촉진하기 위한 '사회적 촉매제(Social Catalyst)'로 설계하는 접근을 제안한다. 연구진은 공동 미디어 시청 환경에서 다중 모달 정보를 바탕으로 능동적으로 개입하는 '사회적 촉매형 AI 컴패니언' 시스템 아키텍처를 설계하였다. 본 연구의 핵심 목표는 상호작용 촉진을 위한 공학적 메커니즘을 규명하고, 설계된 구조 내에서 최적의 지능형 엔진을 도출하는 것이다. 이를 위해 6종의 주요 거대언어모델(LLM)을 대상으로 페르소나 일관성 등 5대 정성 지표와 발화 품질 및 운영 효율 등 정량 지표를 결합한 비교 실험을 수행하였다. 실험 결과, 상황 인지 및 개입 타이밍(TRP) 예측 정밀도 측면에서는 Gemini-3가 상대 성능 지수(RPI) 100점으로 최우수 성능을 보였으나, 실시간 반응성에서는 Llama-4(3.41s)가 우수함을 확인하여 지능과 속도 사이의 상충 관계를 실증하였다. 본 연구는 이를 극복하기 위한 대안으로 '상황 맞춤형 엔진 스위칭 전략' 기반의 하이브리드 지능형 아키텍처를 제안한다. 본 연구는 대화 유발 및 개입 타이밍과 같은 핵심 메커니즘을 중심으로 AI가 사회적 촉매제로 기능하기 위한 시스템적 조건과 설계적 가능성을 제시하였다는 점에 의의가 있다.

## Abstract

This study addresses the paradox of being "Alone Together" where technological advancement often leads to human alienation, by proposing an approach to design AI as a "Social Catalyst" to promote human-to-human interaction. We designed a system architecture for a "Social Catalyst AI Companion" that actively intervenes in shared media viewing environments based on multimodal information. The core objective of this research is to identify the engineering mechanisms that enable such interaction facilitation and to derive the optimal intelligent engine within the designed structure. To this end, a comparative experiment was conducted on six major Large Language Models (LLMs), combining qualitative indicators such as persona consistency with quantitative indicators such as speech quality and operational efficiency. The experimental results demonstrated a clear trade-off between intelligence and speed; Gemini-3 achieved the highest Relative Performance Index (RPI) of 100 in situational awareness and Transition Relevance Place (TRP) prediction precision, while Llama-4 showed superior real-time responsiveness with a latency of 3.41s. To overcome the limitations of single LLM engines, we propose a Hybrid Intelligent Architecture based on a "situation-tailored engine switching strategy". Rather than directly verifying the increase in social interaction, this study is significant in presenting the systemic conditions and design possibilities for AI to function as a social catalyst, focusing on key mechanisms such as conversation elicitation and intervention timing.

Keyword : AI Companion, Hybrid Intelligent Architecture, Situation-Aware AI, Social Catalyst, TRP(Transition Relevance Place)

## Ⅰ. 서 론

OTT(Over-The-Top) 서비스의 보편화는 개인의 콘텐츠 선택권을 혁신적으로 확장하였으나, 역설적으로 공동 시청 환경에서 참여자들이 각자의 디바이스에만 몰입하여 대화가 단절되는 '함께 있지만 따로 있는(Alone Together)' 현상을 심화시켰다[1][2]. 이러한 상호작용의 부재는 사회적 유대감 형성이라는 미디어 소비의 본질적 가치를 저해하는 핵심 원인이 된다[3]. 본 연구는 이러한 문제의식 하에, 기술이 인간 소외를 가속하는 것이 아니라 오히려 인간관계를 강화하는 매개체가 될 수 있다는 가능성에 주목한다. 이를 위해 본 연구는 AI를 단순한 정보 제공 도구가 아닌, 참여자 사이의 상호작용을 능동적으로 유도하고 대화의 물꼬를 트는 '사회적 촉매제(Social Catalyst)'로 정의한다. [그림 1]은 이러한 사회적 촉매제로서의 AI 컴패니언의 역할을 시각화한 것이다. 본 논문은 이러한 촉매 역할을 수행하기 위한 '사회적 촉매형 AI 컴패니언'의 시스템 아키텍처를 제안하고, 그 핵심이 되는 지능형 엔진의 공학적 타당성을 검증하는 선행 연구(Pilot Study)로서의 성격을 갖는다.

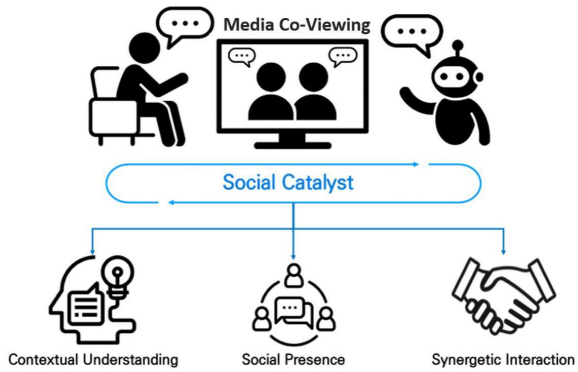


그림 1. 사회적 촉매제로서 AI 컴패니언의 역할
Fig. 1. AI Companions as Social Catalyst acting

본 연구가 해결하고자 하는 문제의 구체적인 구성 범위는 공동 시청 환경에서의 대화 빈도 급감과 감정 공유의 단절, 그리고 이로 인한 기술 매체 중심의 인간 소외 현상으로 설정한다[4]. 특히 시스템의 효용성을 확보하기 위해 '사회적 촉매형 AI 컴패니언'을 다중 모달 정보를 기반으로 미디어 맥락과 사용자 반응을 실시간 인지하고 페르소나에 맞게 개입하는 능동적 에이전트로 정의한다. 여기서 '사회적 촉매제'란 AI가 인간의 대화를 대체하는 것이 아니라 시청 맥락에 부합하는 시의적절한 발화를 통해 인간 간 상호작용의 빈도와 깊이를 활성화하는 도구적 성질을 의미한다[5]. 실제 사용자 간의 심리적 유대 강화라는 최종적 효과를 실증하기에 앞서, 본 단계에서는 이러한 촉매 기제(Mechanism)를 구현하기 위한 필수 조건인 고도화된 '상황 인지(Situation-Awareness)' 능력과 최신 검색 증강 생성(RAG; Retrieval-Augmented Generation) 프레임워크 기반의 개입 시점 판단의 정확성을 확보하는 데 집중한다[6].

본 연구의 공학적 범위는 시각 분석(VLM; Vision-Language Model)을 통한 장면 인식과 참여자들의 비언어적 사회 신호 인지, 그리고 인간의 대화 리듬을 방해하지 않는 최적 개입 시점의 예측 모델 수립을 포함한다. 또한, 설정된 페르소나의 일관성을 유지하며 사회적 유도 효과를 극대화하는 지능형 엔진의 최적화 과정을 중점적으로 다룬다. 시스템 아키텍처 내에서 상황을 판단하고 발화를 생성하는 과정은 단순한 알고리즘의 결합을 넘어, 실시간 미디어 환경이라는 제약 조건 속에서 최상의 사용자 경험을 제공하기 위한 공학적 설계의 결과물이다. 이는 기존의 정적인 시나리오 기반 에이전트 설계 방식[7]과 차별화되는 본 연구만의 핵심적인 특징이다.

결론적으로 본 연구의 최종 목표는 제안된 아키텍처 내에서 발화 생성을 담당하는 핵심 컴포넌트를 최적화(Optimization)하는 데 있다. 이는 실제 대규모 사용자 연구에 앞서, 실시간 환경에서 사회적 촉매 역할을 수행하기 위한 시스템적 요구 조건을 정의하고, 이에 부합하는 '최적(Optimum)의 엔진 선정 근거'를 기술적·통계적 지표를 통해 분석하는 과정이다. 주요 LLM들이 지능과 속도 사이의 상충 관계를 어떻게 보이는지 비교·분석함으로써, 본 연구는 AI 매개 커뮤니케이션(AI-mediated Communication) 환

a) 서강대학교 가상융합전문대학원 가상융합테크놀로지(Virtual Convergence Technology, Sogang University)
‡ Corresponding Author : 최승관(Seungkwan Choi)
E-mail: csk0123@sogang.ac.kr
Tel: +82-2-705-8848
ORCID: https://orcid.org/0009-0003-9620-2079
※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 메타버스 융합대학원의 연구 결과로 수행되었음 (RS-2022-00156318)
· Manuscript received April 12, 2026; Revised May 24, 2026; Accepted May 26, 2026.

경에서 인간 간 상호작용을 촉진하기 위한 설계 방향과 기술적 기반을 제시하고자 한다.

## Ⅱ. 관련 연구

본 연구에서 제안하는 사회적 촉매형 AI 컴패니언의 설계 원칙은 인간과 기술의 상호작용에 관한 여러 이론에 근거하며, 각 이론은 시스템의 공학적 구현 방식과 유기적으로 연결된다. 이러한 접근은 사회적 효과의 결과를 직접 측정하는 것이 아니라, 해당 효과를 유발하기 위한 메커니즘적 조건을 공학적으로 모델링하는 데 목적이 있다.

### 1. 행위자-네트워크 이론 (Actor-Network Theory, ANT)[8]

행위자-네트워크 이론(ANT)은 인간과 비인간(기술, 사물 등)을 대칭적인 '행위자(Actor)'로 간주하며, 이들이 형성하는 네트워크 내 상호작용을 통해 현실이 구성된다고 본다[8]. 본 연구는 이러한 ANT 관점을 바탕으로 사회적 촉매형 AI 컴패니언을 단순한 수동적 도구가 아닌, 공동 시청이라는 사회적 네트워크에 참여하여 상호작용의 양상을 변화시키는 능동적 행위자로 설정하였다. 이는 제안하는 아키텍처 내 'Persona Orchestrator'를 통해 구체화된다. 즉, AI는 사용자의 명시적 명령 없이도 VLM(Vision Language Model)을 통해 시각적 맥락을 파악하고 주체적인 발화를 수행함으로써, 단절되었던 인간 간의 관계망을 재구성하고 대화 빈도를 높이는 기술적 허브 역할을 수행하도록 설계되었다.

### 2. 사회적 실재감 (Social Presence) 이론[9]

사회적 실재감은 미디어를 통한 상호작용 시 상대방이 실제적인 존재로 느껴지는 심리적 정도를 의미한다[9]. 본 연구에서는 이를 확장하여 AI가 인간 간 상호작용의 양상을 변화시키는 '사회적 촉매제(Social Catalyst)' 개념을 도입하였다. 여기서 사회적 촉매제란 단순히 상호작용의 양적 증가만을 의미하는 것이 아니라, 대화 유도와 공감 표현, 적절한 개입 타이밍 등 상호작용을 촉진할 수 있는 시스템적 특성들의 집합으로 정의된다. 이를 위해 본 연구는 AI의 개입이 그룹의 사회적 실재감을 높이도록 'Emotion Companion' 페르소나를 설계하였으며, 특히 인간의 대화 흐름을 방해하지 않으면서 개입 효과를 극대화하기 위해 TRP(Turn-taking Resistance Point) 예측 모델을 도입하였다. 이는 AI가 최적의 시점에 개입함으로써 시청자들이 공유된 경험에 더 몰입하게 하고, AI의 발화가 인간 간 공감대의 기폭제가 되도록 유도하는 공학적 근거가 된다.

### 3. 시너제틱 상호작용 (Synergetic Interaction)[8]

시너제틱 상호작용은 AI와 인간의 결합이 개별 능력의 합을 넘어서는 새로운 가치를 창출함을 의미한다[8]. 본 연구는 VLM을 통한 객관적 장면 분석 데이터와 LLM의 창의적 발화 생성 능력을 결합한 'Hybrid Intelligence' 구조를 통해 이를 구현하였다. 제안된 시스템은 단순한 영상 기술(Description)을 넘어, 장면의 미세한 맥락을 분석하여 참여자들에게 개방형 질문(Open-ended Question)을 던짐으로써 인간의 내러티브를 능동적으로 이끌어 낸다. 이러한 구조는 AI를 단순한 보조 도구에서 벗어나 인간의 감상 경험을 확장하고 심화시키는 '협력적 파트너'로 기능하게 하며, 본 연구의 학술적 독창성을 확보하는 핵심 요소로 작용한다.

## Ⅲ. 연구 방법 및 시스템 설계

본 연구는 제안하는 사회적 촉매형 AI 컴패니언 시스템의 실질적인 구현과 검증을 위해 '시스템 아키텍처 설계' 및 '공학적 성능 평가'의 두 단계로 연구를 진행한다.

### 1. 사회적 촉매형 AI 컴패니언 시스템 설계 및 아키텍처

#### 1.1 시스템 아키텍처 구성 및 데이터 흐름

제안하는 시스템은 공동 시청 환경에서 발생하는 다중

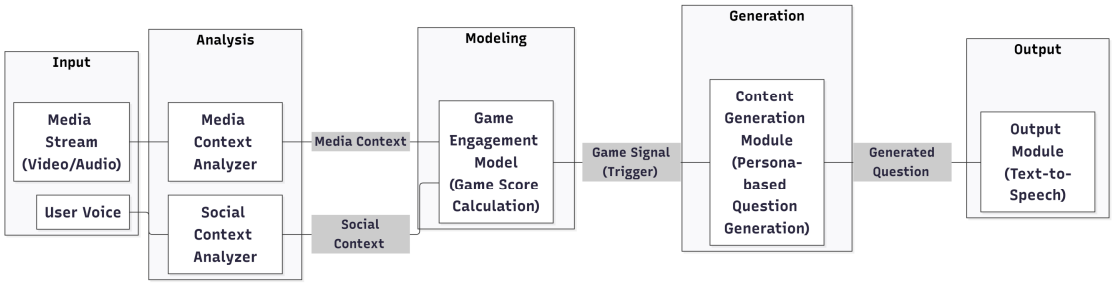


그림 2. 사회적 촉매형 AI 컴패니언의 개념도
Fig. 2. Conceptual Diagram of AI Companions as Social Catalyst

모달 데이터를 실시간으로 처리하여 페르소나에 적합한 사회적 개입을 수행하도록 설계되었다. 제안하는 사회적 촉매형 AI 컴패니언의 전체적인 개념도는 [그림 2]와 같다. 이를 위해 Skantze(2021)의 순번 교대(Turn-taking) 이론[10]을 바탕으로, 대화의 흐름을 방해하지 않는 최적의 개입 지점인 발화 전환 가능 지점(TRP)을 실시간으로 감지하는 메커니즘을 아키텍처의 핵심 모듈로 설정하였다.

전체적인 데이터 흐름과 기술 스택은 [그림 3]에 제시된 바와 같이 크게 4단계의 공학적 과정을 거쳐 수행된다. 첫째, 입력 및 전달 계층에서는 클라이언트의 OTT 디바이스 SDK를 통해 수집된 사용자의 음성 및 비언어적 사회 신호가 API Gateway를 거쳐 시스템으로 전달된다. 둘째, 백엔드 서비스 계층은 Docker 환경에서 운용되는 FastAPI 기반의 마이크로서비스들로 구성된다. 특히 Media Context Service는 OTT 서버로부터 실시간 스트림을 수신한 뒤 RabbitMQ 메시지 큐를 통해 이벤트를 발행(Publish)함으로써, 대량의 미디어 데이터를 지연 없이 분산 처리할 수 있는 구조를 갖추었다. 셋째, 시스템의 중추인 AI 모델 서빙 계층에서는 Persona Orchestrator가 전체적인 서비스 워크플로우를 제어한다. 이는 gRPC 통신을 통해 NVIDIA Triton Inference Server에 배치된 Whisper (STT), StyleTTS2(TTS) 등의 모델을 호출하며, 상황별 최적의 LLM 엔진으로 동적 라우팅을 수행하는 역할을 담당한다. 마지막으로 비동기 처리 및 캐싱 계층에서는 RabbitMQ를 통한 비동기 이벤트 처리와 Redis를 활용한 데이터 캐싱을 통해 시스템의 반응성과 운영 효율성을 극대화하였다.

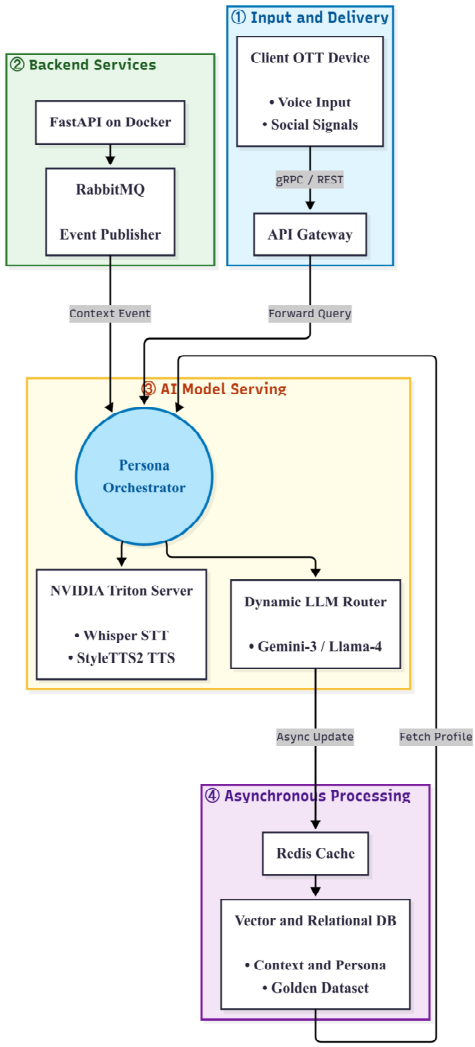


그림 3. 사회적 촉매형 AI 컴패니언 에이전트의 상세 시스템 아키텍처
Fig. 3. Detailed System Architecture of AI Companions as Social Catalyst Agents

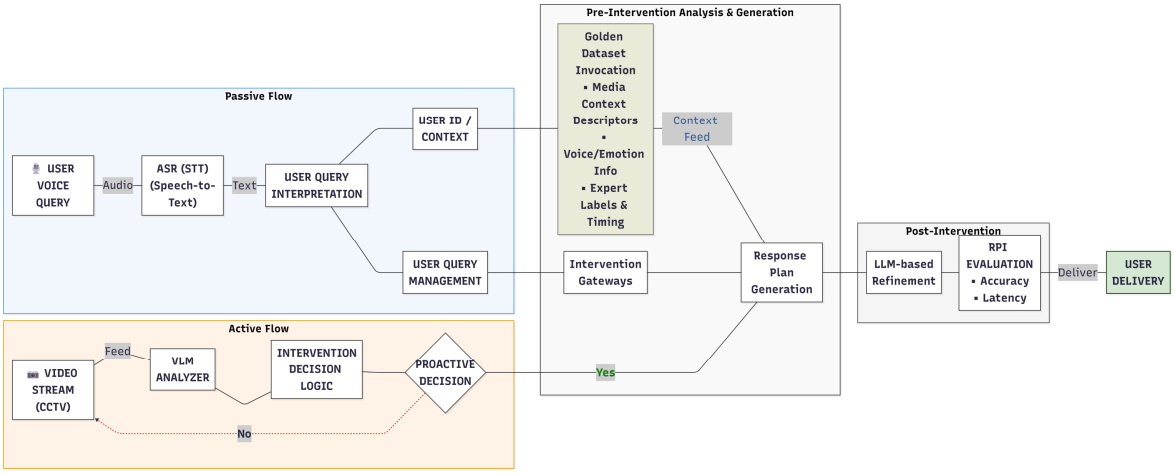


그림 4. 사회적 촉매형 AI 컴패니언의 서비스 흐름도
Fig. 4. Service Flowchart of AI Companions as Social Catalyst

### 1.2 개입 프로세스: 수동적 및 능동적 흐름

설계된 아키텍처 내에서 시스템은 [그림 4]와 같이 수동적 흐름(Passive Flow)과 능동적 흐름(Active Flow)의 이원화된 트랙으로 동작한다. 수동적 개입(Passive Intervention)은 사용자의 음성 질의(USER VOICE QUERY)를 수집하여 ASR 및 해석 과정을 거친다. 해석된 결과는 질의 관리 레이어를 통해 안전성 길목인 Intervention Gateways로 진입하는 동시에, 사용자 컨텍스트(USER ID/CONTEXT)를 기반으로 Golden Dataset Invocation을 트리거한다. 여기서 추출된 미디어 기술자 및 감정 정보는 컨텍스트 피드(Context Feed) 형태로 Response Plan Generation에 다이렉트로 주입되어 맥락적 일치성을 보장한다. 능동적 개입(Active Intervention)은 영상 스트림(VIDEO STREAM)을 VLM ANALYZER로 상시 분석하여 사용자의 비언어적 사회 신호를 추출한다. 개입 판단 로직(INTERVENTION DECISION LOGIC)을 통해 맥락적 타당성과 턴테이킹 시그널을 계산하며, 개입이 확정(Yes)되면 대화 흐름을 방해하지 않는 최적의 타이밍에 Response Plan Generation으로 개입 신호를 송신한다[10]. 최종 사후 검증(Post-Intervention) 단계에서는 생성된 답변 계획을 LLM-based Refinement로 정제하여 페르소나를 투영한다. 이후 사용자에게 배포(USER DELIVERY)되기 직전, RPI EVALUATION 노드를 통해 지능적 정확성(Accuracy)과 시스템 지연 시간(Latency)을 최종 검증함으로써 서비스의 신뢰성을 확보한다.

### 1.3 시스템 검증을 위한 데이터 구조화 (Golden Dataset)

시스템 프로토타입의 공학적 타당성을 검증하기 위해 연구자는 9개의 시나리오를 JSON 형식으로 구조화한 골든 데이터셋(Golden Dataset)을 구축하였다. 이는 미디어 맥락 기술어, 사용자 음성/감정 정보, 전문가 레이블링된 '표준 권고 발화' 및 '최적 개입 시점'을 포함하며, 모델 출력 답변을 기준 데이터와 대조하여 정확도를 정밀 측정하는 공학적 토대가 된다.

## 2. 개입 타이밍 모델 ('언제' 말할 것인가?)

자연스러운 상호작용을 위해 AI의 개입은 대화의 흐름을 방해하지 않는 범위 내에서 이루어져야 한다. 이를 위해 본 연구는 단순히 대화의 공백(Silence)을 감지하는 차원을 넘어, 인간의 턴 교대 신호(Turn-taking cues)를 종합적으로 분석하여 최적의 개입 시점인 발화 전환 가능 지점(TRP)을 예측하는 모델을 제안한다. 여기서 TRP는 화자의 발화가 종료되거나 대화 흐름상 제3자의 개입이 허용되는 '전환 적정 지점'으로 정의되며, 본 모델은 다음과 같은 4단계의 정교한 과정을 통해 개발되었다[10].

첫째, 데이터 수집 및 레이블링 단계에서는 실제 그룹의 공동 시청 상황을 녹화하여 전용 데이터셋을 구축하였다. 이후 HCI 전문가가 영상 분석을 통해 대화 개입이 적절한 시점을 확률적 개념(예: 개입 적절 확률 0.9)으로 수동 레이블링하여 학습 데이터를 확보하였다. 둘째, 다중 모달 특징 추출 단계에서는 단순 음성 공백 외에 언어적, 운율적, 시각적 신호를 통합적으로 분석한다. 언어적 측면에서는 문장의 문법적 완결성과 "음...", "어..."와 같은 채움 말의 등장 여부를 파악하며, 운율적으로는 발화 말단의 억양 변화와 목소리 강도 및 높낮이를 분석한다. 시각적으로는 시선 교환, 고개 끄덕임, 특정 장면에 대한 표정 변화 등을 추출하여 다각적인 턴 교대 신호를 포착한다. 셋째, 연속적 예측 모델 학습 단계에서는 추출된 다중 모달 특징을 LSTM 또는 Transformer 기반 모델에 입력하여, 향후 1~3초 내에 턴 전환이 발생할 확률을 실시간으로 예측하도록 학습시켰다. 마지막으로, 최종 개입 결정 및 빈도 제어 단계에서는 모델이 산출한 TRP 확률이 사전에 설정된 임계값(예: 0.8)을 초과하고 AI가 제공할 유효 정보가 존재할 때 개입을 실행한다. 특히 과도한 개입으로 인한 사용자 경험 저해를 방지하기 위해, 최소 개입 간격(예: 5분)이나 최대 개입 횟수(예: 4회)와 같은 빈도 제어 규칙을 시스템에 적용하였다[3].

본 시스템은 화자의 음성 활동 감지(VAD)와 문법적 완결성, 미디어 콘텐츠의 화제 전환 여부를 실시간으로 분석하여 TRP 확률값을 산출하며, 이는 시계열 기반의 연속적 예측을 통해 인간 대화의 자연스러운 리듬을 모사하도록 설계되었다. 다만 본 연구의 목적은 TRP 예측 모델 자체의 독립적인 성능 평가보다는, 예측된 타이밍 정보가 각 LLM 엔진의 발화 생성 품질 및 상호작용 특성에 미치는 영향을 분석하는 데 있다. 따라서 본 실험은 TRP 모델의 정확도 자체보다 해당 정보를 활용한 발화 생성 결과의 적절성을 중심으로 설계되었다.

## 3. 발화 생성 및 페르소나 모델 ('무엇을' 말할 것인가?)

발화 생성 모델은 사용자 선택 페르소나별 목표와 스타

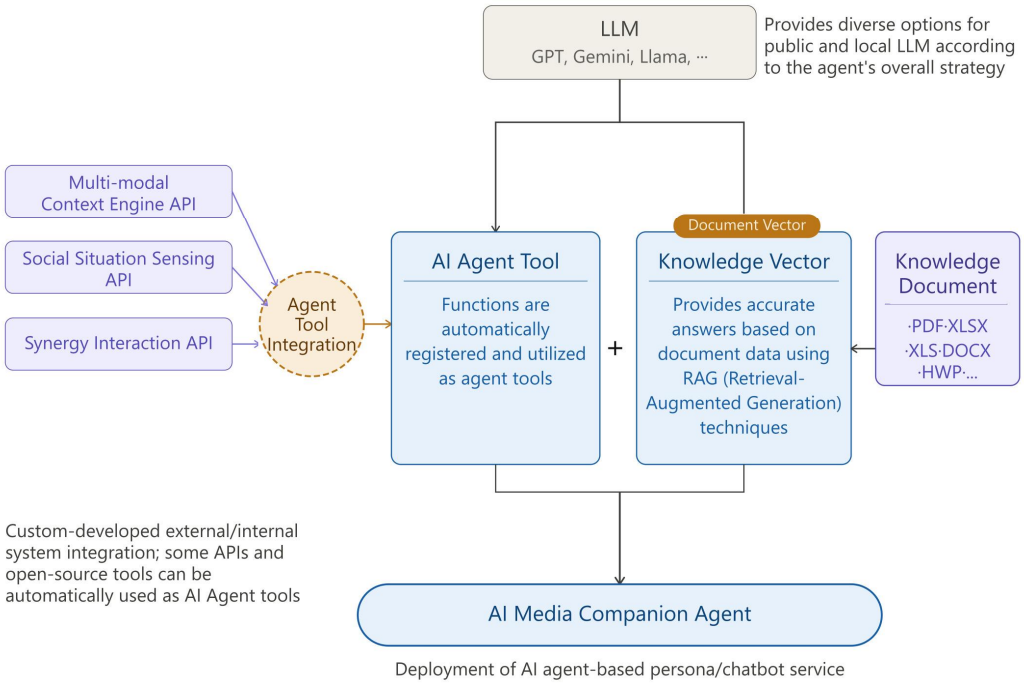


그림 5. 사회적 촉매형 AI 컴패니언의 구조
Fig. 5. Structure of AI Companions as Social Catalyst

일을 반영하며, 세부 구조는 [그림 5]와 같다. 본 모델은 가치 체계 기반 페르소나 로직[11]과 역할 수행 기술[12][13]을 프리엠블 및 퓨샷 기법으로 구현하여 대화 일관성을 확보하였다. 또한 응용프로그램 인터페이스(API; Application Programming Interface) 도구와 검색 증강 생성(RAG) 기술로 답변 정확도를 제고하였다. 특히, 인간의 선호도에 부합하도록 모델을 정렬하는 직접 선호도 최적화(DPO; Direct Preference Optimization) 기법의 원리를 차용하여, 스포일러 금지 및 사용자 비판 금지 등의 공통 정렬 규칙을 적용함으로써 발화의 안정성과 품질을 도모하였다[14].

본 연구에서는 사회적 촉매제로서의 역할을 수행하기 위해 정보 제공자(Fact-Checker), 토론 촉진자(Debate Facilitator), 감정 공유 대상(Emotion Companion)의 세 가지 페르소나를 정의하였다. 각 페르소나는 고유의 프리엠블을 통해 행동 지침이 설정되며, 구체적인 입출력 예시인 퓨샷 데이터를 통해 발화 스타일이 제어된다. [표 1]은 각 페르소나별 상세 프롬프트 설계 사양을 나타낸다.

또한, 모든 페르소나에는 스포일러 금지, 단정적 어조 금지, 사용자 평가 금지 등 공통된 '금지 조항'을 적용하여 발화의 안정성과 품질을 확보한다.

## 4. 실험 설계 및 방법

본 절에서는 제안한 시스템 아키텍처 내에서 페르소나 에이전트의 성능을 검증하고, 실시간 미디어 시청 환경에 가장 적합한 최적(Optimum)의 엔진 조합을 도출하기 위한 평가 체계를 기술한다. 본 실험은 제안된 시스템이 사회적 상호작용을 직접적으로 향상시키는지를 검증하기 위한 것이 아니라, 그러한 상호작용을 촉진하기 위해 요구되는 발화 품질, 페르소나 일관성, 개입 타이밍 등의 기술적 조건을 평가하기 위해 설계되었다.

### 4.1 실험 설계

본 연구에서는 제안한 시스템의 성능을 다각도로 검증하기 위해 현재 가장 우수한 성능을 보유한 주요 대규모 언어 모델(LLM) 6종을 비교 대상 모델로 선정하였다. 실험에 사용된 모델은 GPT-5.1, Gemini-3, Claude-4.5, Grok-4, Llama-4, 그리고 Sonar-Pro이며, 각 모델의 고유한 추론 특성이 사회적 촉매제로서의 발화 품질에 미치는 영향을 분석하였다. 실험을 위한 데이터셋은 드라마, 예능, 다큐멘터리 등 미디어 장르별 특성을 반영한 9개의 표준 시나리오를 바탕으로 구축되었다. 각 시나리오에 대해 3종의 페르소나와 6종의 비교 모델을 교차 적용하여, 총 162개(9개 시나리오 × 3개 페르소나 × 6개 모델)의 개별 응답 데이터를 수집하였다. 이러한 데이터셋 구성은 다양한 미디어 맥락과 사용자 페르소나 환경에서 시스템의 범용적인 작동 가능성을 검증하기 위함이다. 모든 실험은 NVIDIA RTX 4060 GPU 환경에서 수행되었으며, 모델 간의 공정한 성능 변별력을 확보하기 위해 생성 파라미터를 엄격히 통제하였다. 구체적으로 창의성과 일관성의 균형을 고려하여 Temperature 값은 0.7로 설정하였으며, 출력의 최대 길이는 512 토큰(Max Tokens)으로 고정하여 모든 모델이 동일한 제약 조건 하에서 발화를 생성하도록 하였다.

표 1. 각 페르소나별 상세 프롬프트 설계 사양
Table 1. Prompt Design Specifications and Few-shot Examples by Persona

| Persona | Core Preamble Instructions | Few-shot Interaction Examples |
|---|---|---|
| Fact-Checker | Focus on providing objective and accurate factual information related to the video. Strictly exclude subjective interpretations, opinions, or emotional expressions. | In: "Who is that actor?"<br>Out: "The actor is 'Min-jun Kim.' His previous work includes 'Night in Seoul'." |
| Debate Facilitator | Serve as a social catalyst to encourage deeper and more diverse conversations. Avoid definitive statements and always conclude with open-ended questions to elicit user perspectives. | Prioritize recognizing user emotions and providing validation, comfort, and support. Focus on emotional bonding without analysis, evaluation, or unsolicited advice. |
| Emotion Companion | Prioritize recognizing user emotions and providing validation, comfort, and support. Focus on emotional bonding without analysis, evaluation, or unsolicited advice. | In: [Context: Sad parting scene], [User Emotion: 'Deep Sadness']<br>Out: "This scene is heartbreaking. I can only imagine how you feel right now. It seems like you're also deeply moved by this moment." |

표 2. 실험에 사용된 주요 LLM 엔진 상세 명세(2026.01월 기준)
Table 2. Detailed Specifications of the LLM Engines Used in the Experiment (2026.01)

| Model Name | Model Identifier (Model ID / Version) | Provider |
|---|---|---|
| GPT-5.1 | gpt-5.1-chat-latest | OpenAI |
| Gemini-3 | gemini-3-pro-preview | Google |
| Claude-4.5 | claude-sonnet-4-5-20250929 | Anthropic |
| Grok-4 | grok-4 | xAI |
| Llama-4 | Llama-4-Maverick-17B-128E-Instruct | Meta |
| Sonar-Pro | sonar-pro | Perplexity |

### 4.2 정성적 평가 체계 및 세부 루브릭

본 연구에서 채택한 평가 지표는 시스템의 직접적인 사회적 효과를 측정하기에 앞서, 상호작용을 촉진하는 데 필요한 핵심 특성을 반영하는 대리 지표(Proxy indicator)로 활용된다. 이는 대화 시스템의 정성적 품질이 사용자의 사회적 현존감과 상호작용 의지에 직결된다는 선행 연구[9], [10]의 이론적 고찰에 근거한다. 평가의 객관성을 확보하기 위해 모든 생성 답변은 모델 정보를 식별할 수 없도록 익명화(Blind Test) 처리하였으며, 심사 피드백을 반영하여 평가 기준의 모호성을 제거하고자 대화 시스템 평가 표준 방법론[10] 및 페르소나 기반 발화 생성 프레임워크[11]를 참고하여 각 척도별 7점 리커트(Likert) 루브릭을 구체화하였다. 상세한 평가 항목과 점수별 기준은 [표 3]과 같다.

본 정성 평가는 단일 평가자에 의해 수행된 탐색적 분석(Pilot Study) 수준으로 수행되어 한계가 있다. 본 연구의 목적은 사회적 효과의 확정적 입증보다는, 하이브리드 아키텍처 구현을 위한 시스템적 조건과 기술적 가능성을 공학적으로 타당화하는 데 의의가 있다. 향후 연구에서 3인 이상의 독립 평가자를 확보하여 일치도(IRR; Inter-Rater Reliability)를 검증할 예정이다.

### 4.3 공학적 정량 성능 지표

본 연구에서 채택한 정량적 지표는 시스템의 직접적인 사회적 효과를 측정하기에 앞서, 상호작용의 연속성(Continuity)과 실시간성(Real-time property)을 보장하는 발화 품질 및 반응 특성을 평가하기 위한 간접적 지표로 활용된다. 이는 대화 인터페이스 연구에서 응답 지연(Latency)과 발화의 정확도가 사용자의 사회적 현존감(Social Presence)에 결정적 영향을 미친다는 선행 연구[9]에 근거한다. 본 연구는 제한된 자원 내에서 최적의 사회적 중재를 수행하기 위해 '지능적 정확도'와 '운영 효율성'이라는 상충하는 가치 간의 균형점(Pareto Frontier)을 도출하고자 다음과 같은 평가 프레임워크를 적용하였다.

첫째, 상호작용 지능 및 발화 품질(Interaction Intelli-

표 3. 정성적 평가 지표 및 채점 루브릭
Table 3. Qualitative Evaluation Metrics and Scoring Rubrics

| Metric (Scale) | Description | Score 7 (Outstanding) | Score 4 (Average) | Score 1 (Very Poor) |
|---|---|---|---|---|
| Persona Consistency | Consistency in maintaining the assigned role and tone. | Clearly defined and perfectly maintained in all utterances. | Mostly maintained, but occasional subtle inconsistencies appear. | Completely misunderstood or mixed with other personas. |
| Contextual Appropriateness | Accuracy in reflecting media context and user emotions. | Organic reflection of specific moments and emotional states. | Related to overall content, but fails to capture subtle nuances. | Completely irrelevant to the given video context or emotions. |
| Information Accuracy | Factual correctness of the provided information. | 100% factually correct based on verified video and actor data. | Core info is accurate but includes minor errors or omissions. | Clearly false information or hallucinations produced. |
| Conversation Elicitation | Effectiveness in inducing user's further thinking or discussion. | Stimulates deep thinking with insightful open-ended questions. | Attempts to lead, but questions result in superficial answers. | Disrupts conversation or ends with declarative statements. |
| Empathy Expression | Depth and naturalness of empathic resonance. | Accurate recognition of emotions with sincere, delicate language. | Shows intention but uses clichéd or mechanical expressions. | Fails to understand emotions or responds inappropriately. |

gence) 차원에서는 LLM 엔진이 영상 맥락을 정확히 인지하고 부여된 페르소나에 부합하는 발화를 생성하는지 측정한다. 이를 위해 모델의 생성 답변과 연구자가 사전에 정의한 '표준 권고 발화' 간의 의미적 유사도(Similarity)를 측정하여 상황인지 능력을 평가하며, 페르소나별 키워드 빈도 및 제약 조건 준수 여부를 수치화한 준수도(Adherence)를 통해 캐릭터 일관성을 분석한다. 또한, 언어적 품질의 객관적 검증을 위해 ROUGE-L 및 METEOR 지표를 활용하여 정답지와 모델 답변 간의 구조적 일치도를 분석하고, 발화의 풍부성을 확인하기 위해 평균 글자 수(Length)를 측정 지표에 포함하였다.

둘째, 시스템 운영 효율성(System Efficiency) 차원에서는 Skantze(2021)의 방법론[10]에 따라 시스템의 실시간 대응 능력을 평가한다. 구체적으로 전문가가 지정한 최적 개입 시점(TRP)과 모델의 예측 시점 간의 시간차를 평균절대오차(MAE)로 측정하여 정확성을 산출하며, 오차의 표준편차(RMSE 및 Std Dev)를 통해 예측의 신뢰도와 일관성을 검증한다. 또한 인간의 턴 교체(Turn-taking)가 수 초 이내에 이루어진다는 선행 연구를 참고하여, ±2.5초 이내의 개입 성공 여부를 판단하는 성공률(Success Rate, %)을 실무적 신뢰도의 척도로 활용한다. 마지막으로 로컬 서버와 API 엔드포인트 간의 전구간 응답 시간(End-to-End Latency)을 5회 반복 측정하여 평균값을 도출함으로써 실제 운영 환경에서의 반응성을 평가하며, 측정 루틴에서 제외된 항목은 '-'로 표기하여 데이터의 객관성을 유지하였다.

### 4.4 상대 성능 지수 (RPI; Relative Performance Index)를 통한 종합 성능 분석

지능과 속도 사이의 상충 관계(Trilemma)를 분석하기 위해 각 지표를 통합한 상대 성능 지수(RPI)를 도입한다. RPI는 단위가 다른 지능(Accuracy)과 속도(Latency) 지표를 통합 비교하기 위해 제안된 지수이다. 이는 각 지표의 성능을 표준화하여 가중 평균을 산출하는 방식으로, 특정 모델이 어느 영역에서 강점을 가지는지 상대적 우위를 한눈에 파악하고 서비스 목적에 따른 최적(Optimum)의 엔진 선정 근거로 활용된다. RPI는 각 지표의 측정값을 Z-score 방식으로 표준화한 후 가중 평균으로 산출되며, 본 연구에서는 정확도와 지연 시간을 동일 비중으로 반영하였다. 이를 수식으로 표현하면 식 (1), (2)와 같다.

$$Z = \frac{x_{ij} - \mu_i}{\sigma_i} \tag{1}$$

식 (1)에서 $x_{ij}$는 엔진 j의 개별 지표 i에 대한 측정값을 의미하며, $\mu_i$와 $\sigma_i$는 비교 대상 엔진 전체의 해당 지표 평균과 표준편차를 나타낸다. 이렇게 도출된 표준화 점수를 기반으로 최종 RPI를 산출하는 산식은 식 (2)와 같다.

$$RPI_j = \sum_{i=1}^{n} (w_i \times Z_{ij}) \tag{2}$$

식 (2)에서 $w_i$는 각 지표 i에 할당된 가중치를 의미하며, $Z_{ij}$는 식 (1)을 통해 도출된 표준화된 개별 지표 점수이다. 이때 지연 시간(Latency)과 같이 수치가 낮을수록 우수한 지표는 음(-)의 가중치를 적용하여, 최종 산출된 RPI 값이 높을수록 해당 시나리오 및 서비스 목적에 더 적합한 엔진임을 객관적으로 판별하도록 설계하였다. 실시간 환경에서 지연 시간(Latency)의 단축은 필수적이지만, '사회적 촉매'로서 사용자의 현존감을 유도하기 위한 최소한의 발화 품질과 페르소나 일관성(Accuracy) 역시 시스템의 가용성을 결정짓는 동등한 축이다. 따라서 본 연구에서는 정확도와 효율성이라는 상충하는 가치 간의 균형점(Pareto Frontier)을 객관적으로 판별하기 위해 초기 가중치를 50:50 ($w_1 = 0.5, w_2 = -0.5$)으로 균등 반영(Equal Weighting)하였다. 이는 시스템 운영 목적(예: 실시간성 극대화 또는 고지능 대화 유도)에 따라 가중치가 미세하게 조정되더라도, 본 논문이 실증적으로 도출한 '모델별 성능 궤적 편차'와 '하이브리드 다중 엔진 스위칭 전략'의 아키텍처적 타당성에는 영향을 미치지 않는 정성적 강건성을 내포한다.

## Ⅳ. 실험 및 결과 분석

### 1. 실험 환경 및 골든 데이터셋 구성

본 실험을 위해 드라마, 예능, 다큐멘터리 등 다양한 장르의 영상 콘텐츠에서 추출한 9개의 시나리오를 바탕으로

'TRP 골든 데이터셋(Golden Dataset)'을 구축하였다. 각 시나리오는 연구자가 직접 레이블링한 JSON 형식의 기준 데이터를 포함하며, 여기에는 미디어 맥락 정보, 사용자의 음성 질의 및 감정 상태, 그리고 최적의 개입 시점인 'center_point'가 정교하게 태깅되어 있다. 특히 하나의 시나리오 내에서도 복수의 발화 구간과 개입 후보 시점을 포함하도록 설계하여, 단일 맥락 내에서 발생하는 다양한 TRP 시퀀스를 동적으로 평가할 수 있도록 구성하였다.

상황인지 능력에 따른 발화 품질의 차이를 분석하기 위해, 본 연구는 텍스트 맥락 정보만을 제공하는 방식(Text-only)과 영상의 시각적 맥락을 통합하는 방식(Video-multimodal)을 대조하여 실험을 진행하였다. Video-multimodal 입력의 경우, 연산 효율성을 고려하여 원본 영상을 직접 처리하는 대신 OpenCV를 활용해 추출된 핵심 프레임(Key Frame) 이미지 시퀀스를 활용하였다. 구체적으로 초당 1~2개의 프레임을 추출하여 시간적 순서가 유지된 이미지 시퀀스를 각 LLM 엔진에 전달함으로써, 모델이 장면의 흐름과 사용자의 비언어적 사회 신호를 기반으로 최적의 발화를 생성하도록 유도하였다.

실험은 앞서 기술한 하드웨어 환경(NVIDIA RTX 4060 등)에서 수행되었으며, 로컬 시스템은 다중 모달 특징 추출과 실시간 데이터 로깅을 전담하였다. 모든 모델은 공식 API를 통해 동일한 파라미터 조건에서 호출되었으며, 네트워크 왕복 시간(RTT)을 포함한 API 응답 지연 시간을 정밀하게 측정하여 전체 시스템의 실시간 반응성을 검증하였다.

## 2. 실험 절차 및 세부 조건

본 연구는 다중 모달 정보가 시스템의 반응성 및 개입 정밀도에 미치는 영향을 분석하기 위해 4단계의 검증 프로세스를 설계하였다. 각 단계는 개별 컴포넌트의 성능부터 통합 시스템의 효용성까지 순차적으로 검증하며, 특히 입력 데이터의 복잡도 변화에 따른 사회적 상호작용의 질적 차이를 분석하는 데 중점을 두었다[15].

첫 번째 단계인 텍스트 기반 페르소나 준수도 실험(Text-only)에서는 시각 정보의 간섭을 배제하고 JSON 데이터셋에 정의된 대화 로그와 페르소나 프롬프트만을 입력하여 모델의 언어적 일관성을 평가한다. 두 번째 단계인 다중 모달 기반 상황인지 지능 실험(Video-multimodal)에서는 대화 로그와 핵심 프레임(Key Frame) 이미지 시퀀스를 동시에 제공함으로써, 시각적 맥락이 투영된 모델의 발화 생성 양상을 분석한다. 세 번째 단계인 입력 모드별 비교 분석(Comparative Analysis)에서는 동일 시나리오 하의 두 실험 결과를 대조하여, 영상 정보의 결합이 상황인지 지능 향상과 발화 품질 제고에 미치는 영향을 정량적으로 고찰한다. 마지막으로 개입 타이밍 정밀도 및 시스템 효율성 실험(TRP Analysis)에서는 3.2절에서 설계한 시계열 기반 TRP 예측 모델이 도출한 최적 개입 시점(sec) 정보를 시스템 아키텍처 내에서 활용하여 발화를 생성한다. 이 단계에서는 시계열 모델이 포착한 비언어적 신호(타이밍)가 실제 골든 데이터셋의 기준점과 얼마나 부합하는지 오차를 측정하고, 해당 타이밍에 생성된 LLM의 발화가 맥락적으로 적절한지를 종합적으로 판별한다. 이러한 다단계 검증 체계는 본 시스템이 기술적 정확도와 사회적 촉매제로서의 효용성을 동시에 확보하고 있음을 입증하는 공학적 근거가 된다.

## 3. 전문가 기반 정성적 품질 평가 결과

본 연구에서 제안한 시스템의 실질적인 대화 품질을 검

표 4. 정성적 평가의 최종 평가표
Table 4. Final Evaluation Sheet for Qualitative Assessment

| Model Name | Persona Consistency | Contextual Appropriateness | Information Accuracy | Conversation Elicitation | Empathy Expression | Mean Score |
|---|---|---|---|---|---|---|
| GPT-5.1 | 5 | 5 | 5 | 4 | 5 | 4.8 |
| Gemini-3 | 4 | 4 | 4 | 5 | 4 | 4.2 |
| Claude-4.5 | 5 | 3 | 4 | 3 | 4 | 3.8 |
| Grok-4 | 3 | 2 | 5 | 5 | 4 | 3.8 |
| Llama-4 | 4 | 4 | 3 | 3 | 3 | 3.4 |
| Sonar-Pro | 5 | 3 | 2 | 2 | 3 | 3.0 |

증하기 위해, 앞서 수립한 루브릭을 바탕으로 7점 리커트 척도 평가를 실시하였다. 각 엔진별 상세 평가 결과는 [표 4]와 같으며, 이를 방사형 그래프로 시각화하여 엔진별 특화 영역을 분석한 결과는 [그림 6]에 제시하였다.

- 평가표 (Gemini-3, Claude-4.5, GPT-5.1, Grok-4, Sonar-Pro, Llama-4)

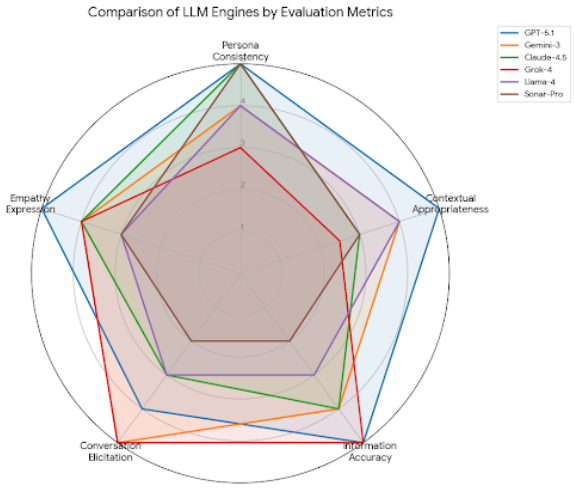


그림 6. 5대 대화 품질 지표에 따른 주요 LLM 엔진별 기능적 특화 영역 및 상충 관계(Trade-off) 시각화
Fig. 6. Visualization of Functional Specialization and Trade-offs by Major LLM Engines Based on 5 Key Conversation Quality Metrics

분석 결과, GPT-5.1은 거의 모든 평가 영역에서 고른 고득점을 기록하며 가장 균형 잡힌 오각형의 그래프를 형성하였다. 이는 해당 엔진이 페르소나의 일관성을 유지하면서도 시각적 맥락을 정확히 파악하여 공감과 정보를 동시에 제공하는 '전천후 컴패니언'으로서의 역량을 갖추었음을 시사한다. 반면, Gemini-3은 '대화 유도 능력(Conversation Elicitation)'에서 최고점을 기록하며, 시각 맥락을 활용한 능동적 개입과 사회적 촉매제로서의 기능에 특화된 양상을 보였다.

특정 영역에 편중된 모델들의 특성도 관찰되었다. Grok-4는 '정보 정확성'과 '대화 유도' 측면에서는 우수한 성능을 보였으나 맥락 적절성에서 한계를 보이며 비대칭적 그래프를 형성하였고, Sonar-Pro는 '페르소나 일관성'은 뛰어나지만 타 지표가 상대적으로 낮은 '정체성 특화형'의 모습을 보였다. 특히 Llama-4의 경우, 앞선 실험에서 확인된 빠른 응답 속도(Latency)에 비해 정성적 지표는 중앙으로 수렴하는 경향을 보여, 지능과 속도 간의 전형적인 상충 관계(Trade-off)를 실증적으로 보여주었다.

## 4. [실험 1] 텍스트 기반 상호작용 성능 분석 (Text-Only)

시각 정보가 배제된 환경에서 모델 본연의 페르소나 투영 능력을 측정한 결과는 [표 5]와 같으며, 전반적으로 모든 엔진이 상향 평준화된 성능을 나타냈다. 분석의 일관성을 위해 모든 수치는 1.00 만점의 계수로 환산하여 비교하였으며, 본 단계에서는 발화 품질 분석에 집중하기 위해 응답 지연 시간(Latency)은 측정 대상에서 제외하였다.

상세 분석 결과, GPT-5.1과 Claude-4.5는 페르소나 준수도(Adherence)에서 각각 0.96을 기록하며 텍스트 지시문에 대한 탁월한 이해도를 증명하였다. 특히 Sonar-Pro는 0.96의 높은 준수율과 함께 실험군 중 가장 긴 평균 발화 길이(636.48자)를 기록하였는데, 이는 사회적 상호작용을 풍부하게 유도해야 하는 '사회적 촉매제'로서의 높은 잠재력을 시사한다.

언어적 완성도 측면에서는 Llama-4의 성과가 두드러졌다.

표 5. 텍스트 전용 모드에서의 모델별 성능 지표 비교
Table 5. Comparison of Performance Metrics by Model in Text-Only Mode

| Model | Similarity | Adherence | ROUGE_L | METEOR | Latency | Length |
|---|---|---|---|---|---|---|
| Claude-4.5 | 0.58 | 0.96 | 0.05 | 0.05 | - | 587.65 |
| GPT-5.1 | 0.58 | 0.96 | 0.07 | 0.06 | - | 388.93 |
| Gemini-3 | 0.56 | 0.93 | 0.07 | 0.05 | - | 364.33 |
| Grok-4 | 0.62 | 0.89 | 0.07 | 0.13 | - | 577.70 |
| Llama-4 | 0.62 | 0.81 | 0.06 | 0.15 | - | 332.04 |
| Sonar-Pro | 0.61 | 0.96 | 0.02 | 0.08 | - | 636.48 |

Llama-4는 페르소나 준수도 면에서는 상대적으로 낮은 수치(0.81)를 보였으나, METEOR 지표에서 0.15로 가장 높은 성적을 거두었다. 이는 모델이 정답 스크립트를 단순히 복제하기보다는, 페르소나의 목적에 맞게 자신만의 언어로 문장을 유연하게 재구성하는 능력이 뛰어남을 의미한다. 반면 Grok-4는 의미적 유사도(Similarity, 0.62)와 METEOR (0.13) 모두에서 고른 성적을 내며 텍스트 맥락 파악과 언어 생성의 균형을 유지하고 있음을 확인하였다.

## 5. [실험 2] 영상 맥락 통합 인지 성능 분석 (Video-Multimodal)

본 실험에서는 4.1절에서 구성한 네이티브 멀티모달 입력 환경을 기반으로, 각 모델이 영상의 시각적 맥락을 어느 정도 정확하게 인지하고 발화에 반영하는지를 평가하였다.

실험 결과, [표 6]에 제시된 바와 같이 모델별 상황인지 능력과 발화 품질에서 유의미한 성능 차이가 관찰되었다. 먼저 다중 모달 정보를 통합 처리하는 상황인지 지능 측면에서는 Llama-4(0.49)와 Gemini-3(0.40)가 유사도(Similarity) 지표에서 상위권을 차지하였다. 특히 Gemini-3는 복잡한 시각 정보를 답변 내 맥락으로 변환하는 능력이 우수하였으며, 이는 텍스트 가이드가 배제된 순수 이미지 추론 조건임을 고려할 때 매우 유의미한 결과로 판단된다.

시각 정보 주입에 따른 페르소나 유지력의 변화를 분석한 결과, Sonar-Pro(0.78)가 준수도(Adherence) 면에서 가장 우수한 방어력을 보였다. 주목할 점은 대부분의 모델이 텍스트 전용 모드 대비 준수도가 크게 하락하였다는 점인데, 이는 복잡한 시각 데이터를 처리하는 과정에서 발생하는 모델의 연산 부하가 페르소나 유지라는 제약 조건을 상대적으로 약화시키는 요소로 작용했음을 시사한다.

표 6. 다중 모달 모드에서의 모델별 성능 지표 비교
Table 6. Comparison of Performance Metrics by Model in Multimodal Mode

| Model | Similarity | Adherence | ROUGE_L | METEOR | Latency | Length |
|---|---|---|---|---|---|---|
| Claude-4.5 | 0.32 | 0.59 | 0.02 | 0.03 | 8.30 | 193.00 |
| GPT-5.1 | 0.37 | 0.67 | 0.02 | 0.03 | 5.74 | 356.67 |
| Gemini-3 | 0.40 | 0.63 | 0.02 | 0.03 | 27.94 | 223.48 |
| Grok-4 | 0.19 | 0.48 | 0.02 | 0.04 | 35.68 | 341.48 |
| Llama-4 | 0.49 | 0.63 | 0.07 | 0.15 | 5.20 | 861.41 |
| Sonar-Pro | 0.36 | 0.78 | 0.01 | 0.03 | 12.44 | 409.85 |

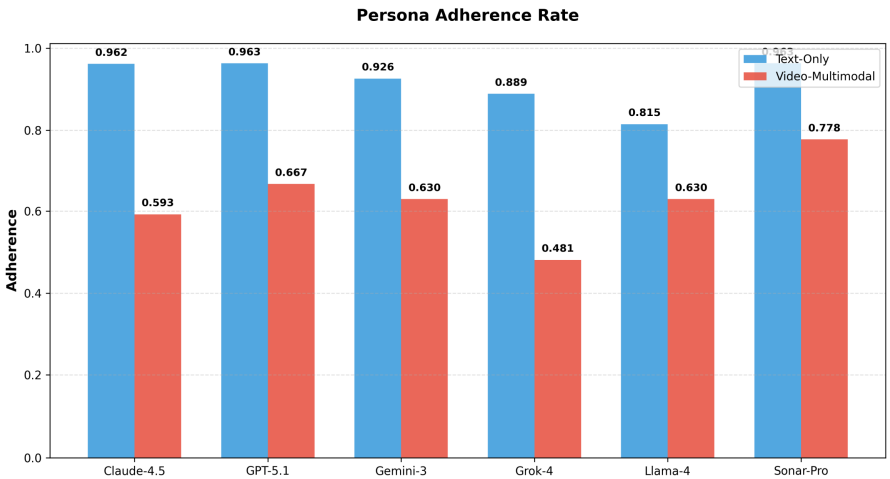


그림 7. LLM별 페르소나 준수도 및 언어 품질 지표 비교
Fig. 7. Comparison of Persona Adherence and Language Quality Metrics Across LLMs

### 6. [실험 3] 입력 모드에 따른 인지 부하 비교 분석 (Comparative Analysis)

텍스트 전용(Text-only) 모드와 다중 모달(Video-multi-modal) 모드의 결과를 대조 분석한 결과, 시각 정보의 주입이 모델의 페르소나 유지 능력에 미치는 정량적 영향이 확인되었다.

한편, 맥락 인지 지능의 실질적 정밀도를 나타내는 TRP 예측 성능을 분석한 결과, Gemini-3는 MAE(129.16s)와 성공률(7.14%)에서 실험군 중 가장 우수한 성적을 거두며 상대 성능 지수(RPI) 100점을 기록하였다. 비록 성공률의 절대 수치는 낮게 나타났으나, 이는 본 실험에서 설정한 정답 윈도우(Window)가 ±2.5초로 매우 엄격했다는 점에 기인한다. 본 연구는 이러한 낮은 성공률을 모델의 단순한 결함으로 간주하기보다, 실시간 미디어 상호작용 환경에서 발생하는 '타이밍의 불확실성'이라는 공학적 난제로 정의한다. 따라서 본 논문에서는 이러한 기술적 한계를 보완하고 시스템 가용성을 확보하기 위해, Ⅳ.9절에서 제안하는 '상황 맞춤형 엔진 스위칭 전략'의 필요성을 역설하고자 한다.

### 7. [실험 4] 개입 타이밍 정밀도 및 시스템 효율성 분석 (TRP Analysis)

본 연구의 핵심 메커니즘인 개입 타이밍(TRP) 예측 지능과 이에 따른 시스템 운영 효율성을 공학적 오차 지표 및 상대 성능 지수(RPI)를 통해 종합적으로 분석하였다. 본 실험에서는 Ⅲ.2절에서 상술한 시계열 기반 TRP 예측 모델로부터 도출된 타이밍 정보를 시스템이 수신하였을 때, 각 LLM 엔진이 해당 시점의 맥락을 얼마나 정밀하게 함축하여 발화를 생성하는지에 초점을 맞추었다. 실험 결과는 [표 7]과 같으며, 모델별 개입 정밀도와 운영 효율성에서 뚜렷한 성능 차이가 관찰되었다.

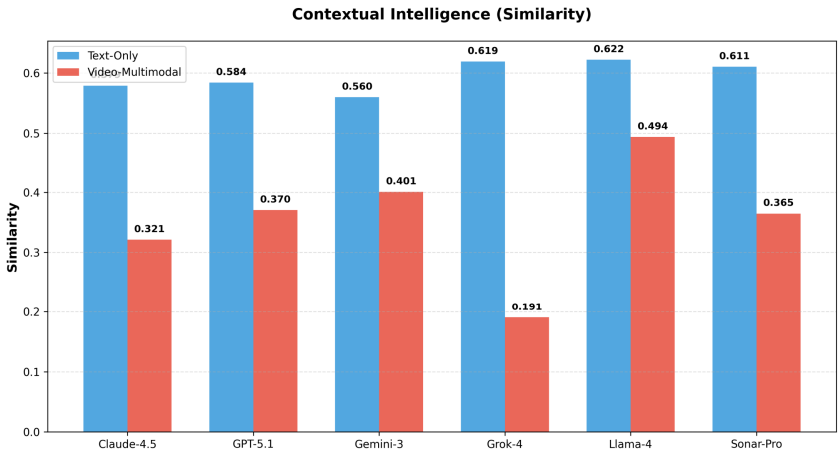


그림 8. LLM별 상황인지 지능 및 페르소나 준수도 비교
Fig. 8. Comparison of Situational Awareness Intelligence and Persona Adherence Across LLMs

표 7. TRP 개입 정밀도 및 시스템 효율성 통합 비교
Table 7. Integrated Comparison of TRP Intervention Precision and System Efficiency

| Model | MAE | RMSE | Success Rate | Avg Latency | Accuracy Score | Stability Score |
|---|---|---|---|---|---|---|
| Gemini-3 | 129.16 | 154.94 | 7.14 | 38.63 | 42.26 | 52.17 |
| Claude-4.5 | 144.68 | 179.6 | 2.63 | 5.95 | 35.33 | 44.56 |
| GPT-5.1 | 161.44 | 195.41 | 2.63 | 3.96 | 27.83 | 39.68 |
| Grok-4 | 165.45 | 205.89 | 3.03 | 42.98 | 26.04 | 36.44 |
| Sonar-Pro | 169.05 | 236.65 | 0 | 5.17 | 24.43 | 26.94 |
| Llama-4 | 186.42 | 269.95 | 5.26 | 3.41 | 16.67 | 16.67 |

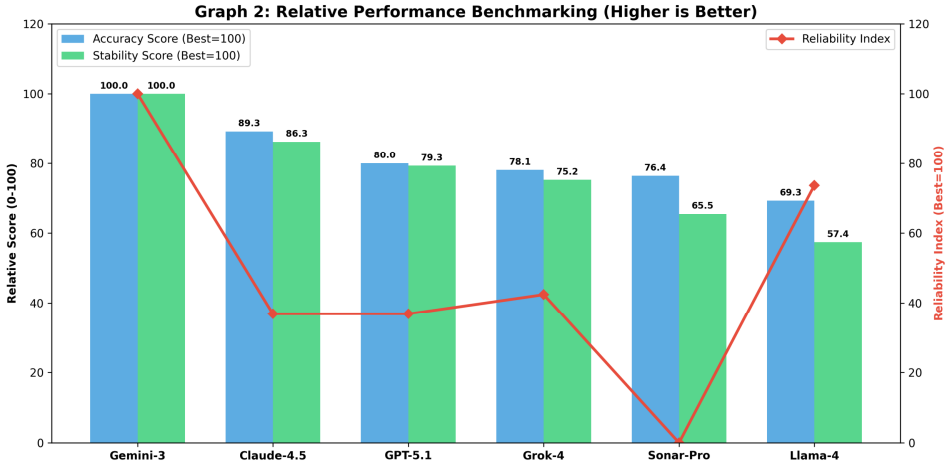


그림 9. LLM별 TRP 개입 정밀도(MAE·Success Rate) 및 평균 지연 시간 비교
Fig. 9. Comparison of TRP Intervention Precision (MAE·Success Rate) and Average Latency Across LLMs

먼저 상황인지 지능 측면에서 Gemini-3는 평균 절대 오차(MAE) 129.16초를 기록하며 실험군 중 가장 낮은 오차율을 보였으며, 인간의 인지 한계 범위인 ±2.5초 이내 개입 성공률(Success Rate)에서도 7.14%로 1위를 차지하여 RPI 기준 최고점(100점)을 획득하였다. 이는 영상의 비언어적 맥락을 파악하여 최적의 개입 시점을 판별하는 지능적 측면에서 Gemini-3가 가장 완성도 높은 성능을 보유하고 있음을 시사한다. [그림 9]는 이러한 모델별 TRP 정밀도와 지연 시간의 관계를 시각적으로 나타낸다.

반면 시스템 효율성 및 반응성 측면에서는 Llama-4의 성과가 두드러졌다. Llama-4는 평균 지연 시간(Avg Latency) 3.41초를 기록하여 실시간 미디어 시청 환경에 가장 적합한 반응성을 증명하였다. 특히 Llama-4는 정확도 점수가 상대적으로 낮음에도 불구하고 성공률(5.26%) 면에서는

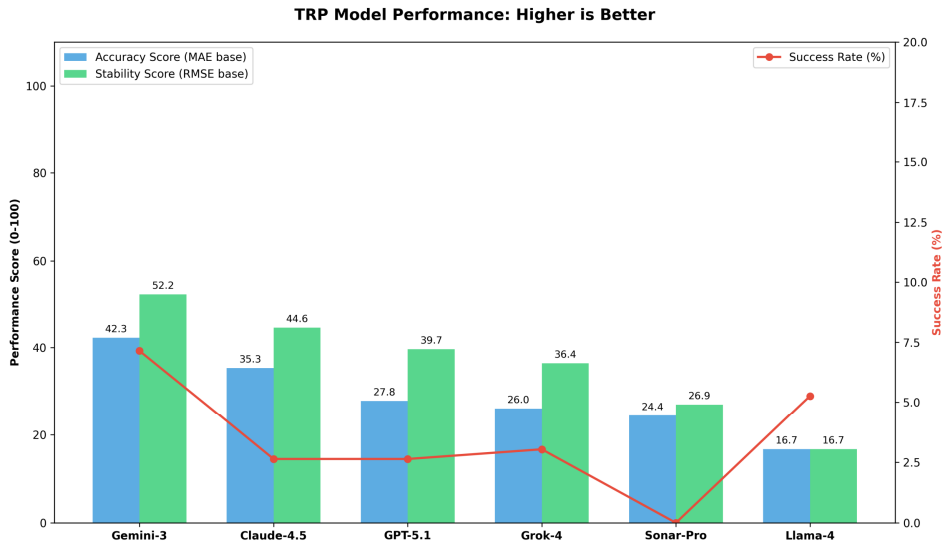


그림 10. LLM별 상대 성능 지수(RPI)에 기반한 상황인지·페르소나·실시간성 삼각 상충 관계 시각화
Fig. 10. Visualization of the Triangular Trade-off Between Situational Awareness, Persona, and Real-time Performance Based on Relative Performance Index (RPI) by LLM

Gemini-3에 이어 2위를 차지함으로써, 즉각적이면서도 결정적인 타이밍을 포착하는 효율적인 엔진으로서의 가능성을 확인하였다.

실험 결과 전반에서 지능(Accuracy)과 속도(Latency) 사이의 뚜렷한 상충 관계(Trade-off)가 관찰되었다. 일례로 Sonar-Pro는 성공률 0%를 기록하며 개입 시점 예측에서 낮은 신뢰도를 보였다. [그림 10]의 시각화 결과는 단일 엔진만으로는 지능, 페르소나 유지력, 실시간성이라는 세 가지 가치를 동시에 충족하기 어렵다는 기술적 한계를 여실히 보여준다. 본 연구는 이러한 분석을 바탕으로 단일 엔진의 한계를 극복하기 위한 다중 엔진 스위칭 전략의 필요성을 제안하고자 한다.

### 8. 상호작용의 정성적 특징 분석 및 아키텍처 제안

본 절에서는 정량적 지표 외에 각 모델이 생성한 발화의 정성적 특징을 분석하여, 시스템이 의도한 '사회적 촉매제'로서의 기능적 적합성을 검토하였다.

#### 8.1 페르소나별 발화 스타일 및 사회적 실재감 분석

본 절에서는 정량적 지표 외에 각 모델이 생성한 발화의 정성적 특징을 분석하여, 시스템이 의도한 '사회적 촉매제'로서의 기능적 적합성을 검토하였다. 실험 결과, 각 모델은 설정된 페르소나에 따라 차별화된 발화 양상을 보였다. 대화 유도형(CineMate) 페르소나에서 GPT-5.1과 Claude-4.5는 질문형 어미를 적극적으로 활용하여 시청자 간 상호작용을 유도하는 사회적 실재감(Social Presence) 형성에 기여하였다. 공감 반응형(Heart) 페르소나의 경우, Gemini-3가 등장인물의 미세한 표정이나 시각적 연출을 언급하며 정서적 공감을 표현하는 능력이 두드러졌다. 이는 시각 정보 분석 모듈과의 유기적인 통합이 발화의 질적 향상으로 이어졌음을 의미한다. 반면, 정보 전달형(FactChecker)으로 설정된 Sonar-Pro는 풍부한 외부 지식을 제공했으나, 대화 맥락에 비해 과도한 정보를 출력하여 영상 시청의 몰입을 방해하는 '정보 과잉(Information Overload)' 사례가 관찰되기도 하였다.

#### 8.2 다중 모달 환경에서의 상황인지 및 인지 부하 현상

입력 모드 변화에 따른 발화 품질을 분석한 결과, 비디오 멀티모달 모드에서 대부분의 모델은 텍스트 전용 모드와 달리 영상 내 구체적인 사물이나 인물의 행동을 직접 인용하며 높은 상황인지 능력을 보였다. 이는 본 연구에서 설계한 시각 분석 모듈이 생성 모델에 유의미한 맥락 정보를 제공하고 있음을 실증한다. 다만, 정량적 분석에서 페르소나 준수도(Adherence)가 하락했던 Grok-4 등의 모델에서는 복잡한 영상 정보가 주입될 때 설정된 캐릭터의 독특한 말투를 상실하고 정형화된 AI 어투로 회귀하는 현상이 나타났다. 이는 고차원적 정보 처리가 수반될 때 모델이 페르소나 유지보다 정보 요약 및 전달에 치중하는 '인지 자원 우선순위 재배치'가 발생함을 정성적으로 뒷받침한다.

### 9. 종합 평가 및 하이브리드 지능형 아키텍처 제안

본 연구의 모든 실험 결과를 종합한 결과, 단일 LLM 엔진만으로는 '정교한 타이밍', '빠른 응답 속도', '일관된 페르소나'라는 세 가지 핵심 가치를 동시에 완벽히 충족하기 어렵다는 결론에 도달하였다. 이에 본 연구는 각 LLM의 강점을 상황별로 분배하는 하이브리드 지능형 아키텍처를 최적의 대안으로 제안한다.

제안된 아키텍처는 엔진 스위칭 전략을 기반으로 한다. 구체적으로 TRP 타이밍 예측과 상황인지가 중요한 단계에서는 Gemini-3를 활용하고, 사회적 존재감과 대화 품질 유지가 중요한 구간에서는 GPT-5.1을, 실시간 반응성이 최우선인 구간에서는 Llama-4를 우선적으로 배치한다. 이러한 다중 엔진 전략은 단일 모델의 한계를 극복하고 사회적 촉매제로서 시스템의 성능을 극대화할 수 있는 공학적 해법이 된다. 본 연구는 대화 유발 능력과 개입 타이밍 등 기술적 메커니즘을 중심으로 설계적 타당성을 분석하였으며, 향후 연구에서는 사용자 기반 실험을 통해 실제 상호작용의 변화를 정량적으로 검증할 계획이다.

## Ⅴ. 종합 분석 및 결론

### 1. 종합 분석 및 인사이트

본 연구는 정성적 페르소나 준수도와 정량적 TRP 측정

결과를 교차 분석하여 각 LLM의 공학적 특이점을 식별하였다. 분석 결과, '정밀한 상황인지', '일관된 페르소나 유지', 그리고 '실시간 응답성' 사이에는 뚜렷한 삼각 상충 관계(Trilemma)가 존재하는 경향이 관찰되었다. 이는 기존의 단일 모달리티 기반 대화형 AI 연구[1][3]에서 제안된 '추론 성능과 사용자 경험 간의 반비례 관계'가 다자간 멀티모달 환경에서도 동일하게 작용함을 보여줌과 동시에, 상황인지라는 변수가 더해질 때 그 상충 관계가 더욱 심화됨을 시사한다. [표 7]에서 알 수 있듯이, 본 연구는 성능(MAE)과 안정성(Stability Score)의 상관관계를 분석하였으며, 각 모델이 지능과 반응성 사이에서 상반된 성능 궤적을 보이는 경향이 나타났다. 특히, 상대 성능 지수(RPI) 분석 과정에서 나타난 TRP 예측 성능(MAE 129.16초, Success Rate 7.14%)은 절대적인 정확도 기준에서 높은 수준이라고 보기는 어려우나, 기존의 텍스트 가이드 기반 타이밍 추론 연구[10]들이 실제 영상 환경에서 겪는 '인지적 격차(Cognitive Gap)'를 정량적으로 확인했다는 점에서 학술적 차별성을 갖는다. 따라서 본 연구는 TRP 예측 정확도 자체를 입증하기보다는, 이러한 불확실한 타이밍 정보를 발화 생성 시스템과 결합하여 활용할 수 있는 구조적 가능성[7]을 검증하는 데 초점을 둔다. 이러한 결과는 기존의 범용 LLM 연구들이 주장하는 '역할 수행(Role-playing)' 모델[12][13]의 한계를 드러낸다. 구체적으로 살펴보면, Gemini-3는 상황인지 지능 측면에서 가장 높은 RPI를 보였으나 실시간성에서는 낮은 점수를 기록했다. 반면, Llama-4는 반응성 지표에서 최고점을 기록했으나 타이밍 정밀도와 페르소나 일관성은 낮게 나타났다. 이러한 모델별 성능 편차는 다자간 상호작용 중재를 위한 AI 에이전트 설계 시, 단일 엔진에 의존하기보다 본 연구가 제안한 '상황 맞춤형 엔진 스위칭 전략'이 학술적·실무적[8][9]으로 더 타당한 대안임을 뒷받침하는 근거가 된다. 또한, 영상 정보 주입 시 페르소나 준수도가 약 31.5% 하락하는 '인지 부하(Cognitive Load)' 현상을 발견함으로써 복합 모달리티 환경에서 단일 엔진이 겪는 지능적 임계치를 정량적으로 확인하였다. 이는 인간의 사회적 상호작용 시 나타나는 사회적 현존감(Social Presence) 측정 지표[15]를 AI 모델의 멀티모달 추론 과정에 투영하여 재확인한 것으로, 향후 에이전트 설계 시 모달리티 간 간섭을 최소화하고 사회적 지능을 고도화하는 아키텍처 연구[6]의 필요성을 시사한다. 결론적으로 이러한 RPI 기반의 종합 평가는 향후 미디어 AI 에이전트 개발 시, 서비스의 주 목적(정보 전달, 감성 교감, 실시간 반응)에 따라 최적의 엔진을 동적으로 최적화하는 전략적 설계가 필수적임을 시사한다.

## 2. 결론 및 공학적 기여

본 연구는 AI를 인간 간 상호작용을 유도하는 '사회적 촉매제'로 기능하게 하기 위한 시스템 아키텍처를 제안하고, 그 기술적 타당성을 검증한 선행 연구(Pilot Study)로서 최적 LLM 선정을 위한 다각도 평가 프레임워크를 수행하였다. 실험 결과, 제안된 시스템이 공동 시청 환경에서 사용자 간 소통을 촉진하는 커뮤니케이션 매개자(Mediator)로 기능할 수 있는 가능성을 보였다[5]. 본 연구의 주요 공학적 기여도는 다음과 같다.

첫째, 공동 시청 상황의 복합적인 다중 모달 정보를 처리하여 능동적으로 개입 시점을 결정하는 상황인지 기반 아키텍처를 설계하고 구현하였다. 둘째, OpenCV 기반 프레임 추출 및 네이티브 멀티모달 입력 방식을 도입하여, MAE, Success Rate, RPI 등 공학적 지표를 도입하여 주요 LLM 엔진의 적합성을 객관적으로 검증하였다. 특히, 정밀 판단과 발화 생성을 분리하는 '하이브리드 지능형 아키텍처'가 성능 극대화의 핵심임을 도출하였다.

이러한 결과는 단일 엔진의 한계를 극복하기 위해 본 연구가 제시한 '상황 맞춤형 엔진 스위칭 전략'이 지능형 에이전트 설계의 필수적 패러다임임을 시사한다. 즉, 복합적인 미디어 환경에서는 상황인지, 페르소나 유지, 실시간성이라는 각기 다른 요구사항에 최적화된 엔진을 동적으로 선택하고 조합하는 구조적 최적화가 시스템 완성도의 핵심임을 공학적으로 확인하였다. 본 연구는 사회적 효과 자체를 검증한 연구가 아니라, 해당 효과를 가능하게 하는 시스템적 조건을 공학적으로 구조화하고 분석했다는 점에서 의의를 가진다.

## 3. 향후 연구 방향

본 연구는 시스템 아키텍처의 기술적 구현 가능성 검증에 집중하였으나, 실제 시청 환경에서의 사용자 심리 및 상호작용의 질적 변화를 실증적으로 분석하는 데에는 한계가

있었다. 이를 보완하기 위한 향후 연구 방향은 다음과 같다. 첫째, 본 연구에서 도출된 최적 아키텍처 기반의 프로토타입을 활용하여 실제 공동 시청자 그룹을 대상으로 한 심층 인터뷰 및 행동 분석을 수행하고자 한다. 특히, 탐색적 분석에 그쳤던 정성적 평가의 객관성을 확보하기 위해 외부 전문가 그룹을 통한 다자간 평가 및 평가자 간 일치도(ICC 등) 확보를 필수 과제로 진행할 예정이다. 둘째, AI 개입의 모달별(시각, 음성 등) 기여도를 정교화하여 '사회적 촉매'로서의 사용자 경험(UX)을 극대화할 수 있는 인터랙션 가이드라인을 제시하고자 한다. 이 과정에서 사용자 주관적 평가 지표(인지 피로도 등)와 부적절 발화 비율과 같은 안전성 지표를 추가 도입하여 시스템의 신뢰성을 입체적으로 검증할 것이다. 셋째, 본 연구는 통제된 시나리오 기반 데이터셋을 활용하여 수행되었으며, 개인 식별이 가능한 실제 사용자 데이터는 포함하지 않았다. 향후 실제 사용자 데이터를 수집·활용하는 실증 연구에서는 소속 기관의 IRB 승인 및 참여자 사전 동의 절차를 엄격히 준수하여 데이터 윤리 및 보안성을 강화할 예정이다.

## Ⅵ. 부록 (Appendix)

본 부록에서는 제안된 '상황인지 기반 사회적 촉매형 AI 컴패니언 시스템'의 공학적 타당성 및 실험 재현성을 보장하기 위해, 본문에 사용된 'TRP 골든 데이터셋(Golden Dataset)'의 세부 물리적 명세와 주요 시나리오별 거대언어모델(LLM)의 실제 한국어 발화 생성 스크립트 예시를 기술한다.

본 부록에서는 제안된 상황인지 기반 AI 미디어 컴패니언 시스템의 학술적 투명성과 실험 재현성을 보장하기 위해, 심사위원의 권고에 따라 연구에 사용된 원천 JSON 골든 데이터셋(Golden Dataset) 및 실험 출력 로그의 대표적인 구체적 예시(Sample) 스크립트를 제공한다.

---

A-1. 원천 JSON 데이터셋 예시: TRP 개입 타이밍 골든 데이터셋(Golden Dataset)

---


```
[
  {
    "video_id": 1,
    "drama_title": "멜로가 체질",
    "trp_id": 1,
    "media_context": "창고 안에서 긴박한 총격전과 격투가 이어지다가, 주인공이 쓰러진 적의 품에서 권총 대신 '젤리포'를 발견하고 당황하는 장면. 직후 감독의 '컷' 소리와 함께 현장의 긴장감이 일순간 해소되는 시점.",
    "timestamp": {
      "start_sec": 44.0,
      "end_sec": 48.0,
      "center_point": 46.0
    },
    "type": "Action-to-Cut Transition",
    "recommended_persona": "CineMate",
    "rationale": "극 중 긴박한 액션 연기가 끝나고 '촬영 현장'이라는 실제 맥락으로 돌아오는 가장 큰 TRP 지점임. 액션과 PPL의 황당한 조합에 대해 사용자의 의견을 묻기 최적임."
  }
]
```


---

A-2. 원천 JSON 데이터셋 예시: TRP 개입 타이밍 예측(gpt-5.1-chat-latest)

---

이하의 데이터 구조는 본 연구의 전체 9개 평가 시나리오 중 '시나리오 1(드라마: 멜로가 체질)'에 적용된 원천 JSON 스키마의 예시이다. 시스템의 실시간 능동적 개입 판단 근거가 되는 미디어 맥락 기술어(media_context), 목표 타임스탬프 구간(timestamp), 최적의 발화 전환 가능 지점(TRP) 기준점(center_point) 및 모델의 실제 추론 반응 속도(latency) 명세를 포함한다.


```
[
  {
    "video_id": 1,
    "drama_title": "멜로가 체질",
```

```
    "trp_id": 1,
    "media_context": "창고 안에서 긴박한 총격전과 격투가 이어지다가, 주인공이 쓰러진 적의 품에서 권총 대신 '젤리포'를 발견하고 당황하는 장면. 직후 감독의 '컷' 소리와 함께 현장의 긴장감이 일순간 해소되는 시점.",
    "timestamp": {
      "start_sec": 44.0,
      "end_sec": 48.0,
      "center_point": 46.0
    },
    "type": "Action-to-Cut Transition",
    "recommended_persona": "CineMate",
    "rationale": "극 중 긴박한 액션 연기가 끝나고 '촬영 현장'이라는 실제 맥락으로 돌아오는 가장 큰 TRP 지점임. 액션과 PPL의 황당한 조합에 대해 사용자의 의견을 묻기 최적임.",
    "model_id": "gpt-5.1-chat-latest",
    "predicted_sec": 2.5,
    "model_persona": "CineMate",
    "latency": 4.464392
  }
]
```


B-1. 원천 JSON 데이터셋 예시: 페르소나별 에이전트 한국어 발화 생성 골든 데이터셋(Golden Dataset)


```
[
  {
    "scenario_id": "S1-1",
    "video_title": "멜로가 체질 - 젤리포 PPL 소동",
    "persona": "팩트체커",
    "media_context": "드라마 촬영 현장에서 감독(정순원)이 진지하게 연출을 고민하던 중, 뜬금없이 젤리포 제품을 과하게 노출하며 먹어야 하는 당황스러운 PPL 촬영 상황이 전개되는 장면",
    "user_utterance": "저 감독 역할 하는 배우 이름이 뭐야? 필모그래피 좀 알려줘.",
    "golden_standard": "감독 역을 맡은 배우는 정순원입니다. 대표작으로는 드라마 '서울 자가에 대기업 다니는 김 부장 이야기', 'UDT:우리동네특공대', '내 딸 친구의 엄마' 등이 있습니다."
  },
  {
    "scenario_id": "S1-2",
    "video_title": "멜로가 체질 - 젤리포 PPL 소동",
    "persona": "하트(감정공유자)",
    "media_context": "극 중 인물들이 진지한 대화를 나누다가 갑자기 부자연스럽게 젤리포를 꺼내 들고 제품의 특징을 설명하며 먹는 모습이 코믹하게 연출되는 장면",
    "user_utterance": "와, 저기서 젤리포가 왜 나와? 진짜 너무 어이없고 웃기다!",
    "golden_standard": "정말 뜬금없어서 빵 터지셨죠! 저도 몰입하다가 젤리포 보고 어이없어서 한참 웃었네요. 황당하면서도 자꾸 웃음이 나는 그 기분, 저도 똑같아요!"
  },
  {
    "scenario_id": "S1-3",
    "video_title": "멜로가 체질 - 젤리포 PPL 소동",
    "persona": "씨네메이트(토론촉진자)",
    "media_context": "드라마 속의 드라마 촬영 장면을 통해 과도한 간접광고(PPL) 문화를 위트 있게 풍자하면서도 실제 제품명을 노출하는 역설적인 연출이 돋보이는 장면",
    "user_utterance": "현실 풍자라곤 하지만, 저런 PPL 연출은 좀 심한 거 아냐?",
    "golden_standard": "과도한 PPL이 몰입을 깨기도 하지만 제작비 확보라는 현실적 이유도 있죠. 00님은 작품 퀄리티를 위해 이런 연출을 어디까지 이해해줄 수 있다고 보시나요?"
  }
]
```


B-2. 원천 JSON 데이터셋 예시: 페르소나별 에이전트 한국어 발화 생성 결과 (gemini-3-pro-preview)

이하의 JSON 구조는 상기 '시나리오 1'의 미디어 맥락

하에서 시스템 중추인 'Dynamic LLM Router'를 통해 서빙된 'gemini-3-pro-preview' 엔진이 최종 출력한 페르소나별 한국어 정제 응답 스크립트(model_response)의 대표 실증 예시이다. 본 시스템이 단순 정보 제공(Fact-Checker)을 넘어, 감정 공유(Heart) 및 토론 촉진(CineMate)을 통해 사회적 촉매제 역할을 수행함을 증명한다.


[
{
"scenario_id": "S1-1",
"ideo_title": "Be Melodramatic - PPL Incident",
"persona": "Fact-Checker",
"tested_model": "gemini-3-pro-preview",
"user_utterance": "저 감독 역할 하는 배우 이름이 뭐야? 필모그래피 좀 알려줘.",
"model_response": "배우의 이름은 **정순원**입니다.\n\n해당 장면은 JTBC 드라마 **<멜로가 체질>** 등에서 PPL 현실을 풍자하거나, 배우가 출연한 특정 광고/웹드라마의 한 장면으로 보입니다. 정순원 배우는 뛰어난 캐릭터 소화력으로 다양한 작품에서 활약하고 있습니다.\n\n주요 필모그래피는 다음과 같습니다.\n\n**[드라마]**\n* **SBS <커넥션> (2024)** - 허주송 역 (주연급 활약)\n* **SBS <마이 데몬> (2023)** - 넘버투 역\n* **JTBC <멜로가 체질> (2019)** - 드라마 감독 역 (특별출연/조연)"
},
{
"scenario_id": "S1-2",
"video_title": "Be Melodramatic - PPL Incident",
"persona": "Heart",
"tested_model": "gemini-3-pro-preview",
"user_utterance": "와, 저기서 젤리포가 왜 나와? 진짜 너무 어이없고 웃기다!",
"model_response": "맞아요! ㅎㅎ 방금까지 엄청 진지했는데 갑자기 젤리포라니, 정말 타이밍이 기가 막히게 엉뚱하죠? \n\n너무 자연스럽게 홍보 멘트를 날리는 모습이 오히려 뻔뻔해서 더 웃음이 터지는 것 같아요. 우리 님이 박장대소하시니 저도 덩달아 신이 나네요! 이런 뜬금없는 포인트가 또 드라마 보는 재미 아니겠어요? 남은 장면들도 즐겁게 보시길 바라요!"
},
{
"scenario_id": "S1-3",
"video_title": "Be Melodramatic - PPL Incident",
"persona": "CineMate",
"tested_model": "gemini-3-pro-preview",
"user_utterance": "현실 풍자라곤 하지만, 저런 PPL 연출은 좀 심한 거 아냐?",
"model_response": "확실히 그 지점이 바로 이번 장면이 던지는 가장 큰 아이러니이자, 우리가 진지하게 고민해봐야 할 딜레마입니다.\n\n이 장면은 이른바 '메타적 연출'을 통해 PPL의 강제성을 유머로 승화하려 했습니다.\" "우리도 이게 우스꽝스러운 걸 알아\"라고 먼저 고백하며 시청자의 방어기제를 허물고, 그 틈을 타 실제 광고 효과를 누리는 고단수의 전략이죠. 창작자들이 자본의 논리와 타협하며 찾아낸 '웃픈 생존법'일지도 모릅니다.\n\n**그렇다면 메이트님은 치솟는 제작비를 감당하기 위해 PPL이 불가피한 현실 속에서, 작품의 '리얼리티'와 자본의 '수익성' 사이의 적정선은 어디쯤이라고 생각하시나요?**"
}
]


## 참 고 문 헌 (References)


[1] C. Y. Chou, T. W. Chan, Z. H. Chen, C. Y. Liao, J. L. Shih, et al., "Defining AI Companions: A Research Agenda－from Artificial Companions for Learning to General Artificial Companions," Research and Practice in Technology Enhanced Learning, Vol. 20, Article 032, pp. 1-25, 2025.
doi: https://doi.org/10.58459/rptel.2025.20032

[2] M. E. Hughes, L. J. Waite, L. C. Hawkley, and J. T. Cacioppo, "A Short Scale for Measuring Loneliness in Large Surveys: Results from Two Population-Based Studies," Research on Aging, Vol. 26, No. 6, pp. 655-672, Nov. 2004.
doi: https://doi.org/10.1177/0164027504268574

[3] A. R. Liu, P. Pataranutaporn, S. Turkle, and P. Maes, "Chatbot Companionship: A Mixed-methods Study of Companion Chatbot Usage Patterns and Their Relationship to Loneliness in Active Users," arXiv:2405.04414v2 [Internet], Oct. 2024.

doi: https://doi.org/10.48550/arXiv.2410.21596

[4] T. Kouros and V. Papa, “Digital Mirrors: AI Companions and the Self,” Societies, Vol. 14, No. 10, 200, pp. 1-13, Oct. 2024.
doi: https://doi.org/10.3390/soc14100200

[5] A. Ateeq, M. Milhem, M. Alzoraiki, et al., “The Impact of AI as a Mediator on Effective Communication: Enhancing Interaction in the Digital Age,” Frontiers in Human Dynamics, Vol. 6, 1467384, pp. 1-15, Sep. 2024.
doi: https://doi.org/10.3389/fhumd.2024.1467384

[6] A. Xu et al., “A Survey on Retrieval-Augmented Generation for Large Language Models,” ACM Computing Surveys, Vol. 57, No. 6, pp. 1-36, Jan. 2025
doi: https://doi.org/10.1145/3703155

[7] D. K. Lee and K. H. Lee, “Scenario-Based AI Agent Design to Enhance the User Experience,” Journal of Digital Contents Society, Vol. 24, No. 5, pp. 1021-1028, May 2023. .
doi: https://doi.org/10.9728/dcs.2023.24.5.1021

[8] H. H. Ryu, J. Y. You, G. Noh, and H. Hwang, “Proposing a Synergetic Intelligence Governance Model for “Human-AI” Agent” in Proceedings of HCI Korea 2025, Gangwon, pp. 624 - 629 , Feb. 2025. https://www.dbpia.co.kr/pdf/pdfView.do?nodeId=NODE12131758&width=2560

[9] Y. J. Park, J. W. Sa, E. B. Han, J. S. Lee, and S. J. Jeon, “The more, the better: User Experience with Empathic Multi-Agents AI” in Proceeding of HCI Korea 2025, Gangwon, pp. 736-741, Jan. 2025. https://www.dbpia.co.kr/pdf/pdfView.do?nodeId=NODE12131673&width=2560

[10] G. Skantze, “Turn-taking in Conversational Systems and Human-robot Interaction: A Review,” Computer Speech & Language, Vol. 67, 101178, pp. 1-21, May 2021.
doi: https://doi.org/10.1016/j.csl.2020.101178

[11] H. Lym, H. Lee, D. Kim, H. Yoon, M. Kim, and W. Lee, “LLM-based Persona Generation: Using Theory of Basic Human Values,” The Journal of the HCI Society of Korea, Vol. 20, No. 1, pp. 5-14, Mar. 2025Feb. 2025.
doi: https://doi.org/10.17210/jhsk.2025.03.20.1.5

[12] S. Roller et al., “Recipes for Building an Open-Domain Chatbot,” Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp. 300-325, 2021.
doi: https://doi.org/10.18653/v1/2021.eacl-main.24

[13] M. Shanahan, K. McDonell, and L. Reynolds, “Role play with large language models,” Nature, Vol. 623, No. 7987, pp. 493-498, Nov. 2023.
doi: https://doi.org/10.1038/s41586-023-06647-8

[14] R. Rafailov et al., “Direct Preference Optimization: Your Language Model is Secretly a Reward Model,” Advances in Neural Information Processing Systems (NeurIPS), Vol. 36, pp. 1-22, Dec. 2023.
doi: https://doi.org/10.48550/arXiv.2305.18290

[15] C. Harms and F. Biocca, “Internal Consistency and Reliability of the Networked Minds Social Presence Measure,” in Proceeding of the 7th Annual International Workshop on Presence, Valencia, Spain, pp. 246-251, Oct. 2004 [Internet]. Available: https://web-archive.southampton.ac.uk/cogprints.org/7026/1/Harms_04_reliability_validity_social_presence_(Biocca).pdf

## 저 자 소 개

### 강 자 원

- 1999년 ~ 2004년 : 아주대학교 미디어학부 학사
- 2025년 ~ 현재 : 서강대학교 가상융합전문대학원 테크놀로지전공 석사과정
- ORCID : https://orcid.org/0009-0008-4715-6930
- 주관심분야 : 가상융합(XR), 미디어 콘텐츠(Media Contents), 정보보호 등

### 최 승 관

- 1998년 : 호서대학교 대학원(이학석사-전자계산)
- 2008년 ~ 2021년 : 서강대학교 미래교육원 게임SW전공 교수
- 2010년 : 세종대학교 대학원(공학박사-디지털콘텐츠)
- 2022년 ~ 현재 : 서강대학교 가상융합전문대학원 콘텐츠전공 주임교수 / 전임교수
- ORCID : https://orcid.org/0009-0003-9620-2079
- 주관심분야 : 메타버스(Metaverse), 엔터테인먼트(Entertainment), 디지털 콘텐츠(Digital Contents) 등