

특집논문 (Special Paper)

방송공학회논문지 제31권 제5호, 2026년 9월 (JBE Vol.31, No.5, September 2026)

<https://doi.org/10.5909/JBE.2026.31.5.830>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

AI 편향성 인지를 위한 다감각 공간 미디어 시스템 아키텍처: 성찰적 UX 기반 인터랙션 설계

강 자 원^{a)}, 최 승 관^{a)*}

Design of a Multimodal Spatial Media System for AI Bias Awareness: A Reflective UX Approach

Jawon Kang^{a)} and Seungkwan Choi^{a)*}

요 약

최근 생성형 AI의 확산과 함께 시스템의 편향성을 무비판적으로 수용하는 자동화 편향이 중요한 과제로 논의되고 있다. 이에 본 연구는 매끄러운 HCI 패러다임에서 벗어나, AI 데이터의 불완전성을 다감각 피드백으로 시공간에 드러내는 참조 아키텍처 <프리즘 AI>를 제안한다. 제안 모델은 AI 지표의 변화를 물리적 거동 변수(*Jerk*), 음향적 불협화도(*Tritone*), 광학 자극(*Flicker*) 등 공간 변수로 매핑하는 ‘감각적 마찰’ 메커니즘을 중심으로 구성된다. 아키텍처의 실현 가능성을 탐색하기 위해 일반 사용자 35인(N=35)을 대상으로 가상 시뮬레이션 기반의 탐색적 지각 평가를 수행했다. 데이터 분석 결과, 지표 전반에서 조건별 평균 스코어의 양적 향상 경향성이 관찰되었으며, 성찰적 UX에 대한 사용자의 초기 지각 경향성을 탐색하였다. 본 참조 모델은 초기 탐색적 평가 단계의 특성상 일반화된 인과관계 검증에는 한계를 가지지만, 향후 성찰적 AI 인터랙션 구현을 위한 다학제적 후속 실증 연구가 참조할 수 있는 개념적·공학적 설계 청사진을 제시했다는 점에서 학술적 의의를 지닌다.

Abstract

As generative AI becomes rapidly proliferated, automation bias—where users uncritically accept system outputs—has emerged as a critical challenge in HCI. To address this, this study departs from the conventional seamless interaction paradigm and proposes “Prism AI,” a conceptual reference architecture that intentionally exposes AI data imperfections through spatiotemporal multimodal feedback. The core of the proposed model is a “Sensory Friction” mechanism that maps real-time AI uncertainty indicators onto physical spatial variables, including kinetic jerk, acoustic tritone dissonance, and optical flicker. To evaluate the architectural feasibility and observe user perceptual tendencies, we conducted an exploratory perceptual evaluation using high-fidelity video artifacts with 35 generative AI users (N=35). Data analysis revealed positive quantitative trends in mean scores across all perceptual metrics, suggesting preliminary perceptual tendencies toward reflective UX. Although this reference model has limitations in validating rigorous statistical significance and generalized causal relationships due to its exploratory nature, it delivers a unique academic contribution by providing a conceptual and engineering design blueprint that multi-disciplinary follow-up empirical studies can readily adopt as a reference archetype.

Keyword : Reflective UX, Immersive Interaction, AI Bias Visualization, Data-to-Sensory Mapping, Cognitive Walkthrough

Copyright © 2026 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

현대 미디어 및 사용자 경험(UX) 패러다임은 사용자에게 단절 없는 몰입과 편의성을 제공하고자 시스템의 기술적 복잡성을 은폐하는 ‘매끄러운 디자인(Seamless Design)’을 지향해 왔다. 내부 연산이나 네트워크 지연 등의 이음새(Seam)를 감추고 결과물만 직관적으로 전달하는 방식은 인터페이스 마찰을 줄여 사용자 만족도를 높였다. 그러나 생성형 AI(Generative AI)가 고차원적 의사결정에 깊숙이 개입하는 현시점에서, 이러한 기술적 은닉성은 알고리즘 내부에 잠재된 사회적 편향성(Bias), 유해성(Toxicity), 블랙박스적 불완전성을 은폐하는 부작용을 낳는다. 시스템이 중립적이라는 환상은 사용자의 비판적 사고를 저해하며, 결과적으로 자동화 편향(Automation Bias)과 기술 예측을 초래한다^[1].

최근 실증 연구들은 AI의 무비판적 수용이 인간의 인지적 제동력을 둔화시키고 고차원적 사유 능력을 감소시킬 수 있음을 경고한다^[2]. AI가 즉각적이고 매끄러운 답변을 제공할수록 인간이 문제를 분석하는 필수적 지적 훈련 과정(Desirable Difficulties)을 우회하기 때문이다. 이러한 인지적 나태(Cognitive Sloth)를 끊기 위해서는 시스템의 한계와 판단 근거를 스스로 사유하도록 유도하는 인터페이스 차원의 개입, 즉 ‘유익한 마찰(Useful Friction)’ 설계가 시급하다. 기존 설명 가능한 AI(XAI) 분야에서도 수치적 지표나 대시보드를 제공하는 연구가 진행되었으나(Ahn et al., 2024)^[3], 이러한 2D 시각화는 텍스트·그래프 중심의 정적 정보 전달에 머물러 일반 사용자의 인지적 각성을 이끌어내기 어렵다^[4]. 따라서 거대 언어 모델(LLM)이 내부적으로 출력하는 추상적인 알고리즘적 한계 수치(편향도, 유해성, 모델 불확실성 엔트로피 등)를 신체적으로 체감

(Physical Embodiment)할 수 있는 새로운 인터랙션 미디어 프레임워크가 요구된다.

이에 본 연구는 기술의 이음새를 의도적으로 노출·자극하는 실감형 연출을 통해 사용자가 AI의 불완전성을 감각적으로 인지하도록 유도하는 상위 수준의 인터랙션 방법론 및 디자인 개념 체계로서 ‘성찰적 UX(Reflective UX)’ 프레임워크를 제안한다. 또한, 이 개념적 프레임워크를 실감형 온·오프라인 스마트 공간 환경에 실시간 데이터 스트림과 수식으로 실현하는 기술 참조 아키텍처(Reference Architecture)로서 <프리즘 AI(Prism AI)>를 설계한다. 제안 모델은 알고리즘의 윤리적 데이터 수치(편향도, 유해성, 불확실성)를 키네틱 상들리에의 물리적 운동량(기울기, Jerk)과 청각적 불협화음, 조명 플리커링으로 실시간 변환하는 데이터-감각 매핑 로직(Mapping Logic)을 핵심으로 한다.

본 연구는 제안 모델의 타당성을 검증하기 위해 가상 시나리오 분석(Virtual Scenario Analysis)과 인지적 워크스루(Cognitive Walkthrough) 방법론을 채택한다. 이를 통해 복잡한 정량 지표가 다감각 마찰로 변전되는 데이터 파이프라인의 공학적 무결성을 진단하고, 해당 자극이 AI 출력에 대한 사용자의 재검토 가능성을 지원하는 설계 조건을 탐색한다. 궁극적으로는 다양한 미디어 도메인에서 기술 윤리 공간 인프라 및 소프트웨어를 설계할 때 원형으로 참조할 수 있는 기술 참조 아키텍처(Reference Architecture)를 제시하고자 한다^[5].

II. 관련 연구 및 이론적 배경

본 장에서는 제안하는 성찰적 UX(Reflective UX) 기반의 다감각 공간 미디어 시스템 아키텍처를 학술적으로 규명하고 연역적으로 검증하기 위해 세 가지 차원의 이론적 토대를 고찰한다. 첫째, 초기 ‘사라지는 기술(Disappearing Computer)’의 철학에서 비롯되어 현대의 인공지능 기반 앰비언트 인텔리전스(Ambient Intelligence)로 진화한 미디어 환경 속에서, 기술의 복잡성과 한계를 다루는 두 가지 대조적 패러다임인 ‘매끄러운 디자인(Seamless Design)’과 ‘이음새 있는 디자인(Seamful Design)’의 개념적 변천을 분석

a) 서강대학교 가상융합전대학원 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 융합 대학원의 연구 결과로 수행되었음 (RS-2022-00156318).

· Manuscript received June 15, 2026; Revised July 14, 2026; Accepted July 14, 2026.

한다. 둘째, 생성형 AI의 일상적 활용이 인간의 고차원적 인지 체계에 미치는 부작용인 ‘인지적 부채(Cognitive Debt)’의 실증적 메커니즘을 살펴보고, 이를 완화하기 위한 개입 장치로서 ‘유익한 마찰(Useful Friction)’과 성찰적 UX의 심리적·인지적 효용성을 검토한다. 마지막으로, 대규모 언어 모델에서 발생하는 고정관념화나 필터 버블 등의 ‘알고리즘적 위해(Algorithmic Harms)’를 시각화해 온 기존 XAI 연구 동향을 파악하고, 스크린 기반 2D 인터페이스의 한계를 극복하기 위한 대안으로서 ‘데이터 물리화(Data Physicalization)’의 미디어 공학적 당위성을 정립함으로써, 본 설계 아키텍처의 연역적 분석을 위한 이론적 기준을 수립하고자 한다.

1. 매끄러운 디자인(Seamless)과 이음새 있는 디자인(Seamful)의 패러다임 변천

초기 유비쿼터스 컴퓨팅(Ubiquitous Computing)에서 출발하여 현대의 앰비언트 인텔리전스(Ambient Intelligence) 및 가상 융합 환경으로 진화한 인간-컴퓨터 인터랙션(HCI) 연구에서, 마크 와이저(Mark Weiser)가 제시한 핵심 가치는 기술이 인간의 일상 속에 자연스럽게 스며들어 의식되지 않는 ‘사라지는 기술(Disappearing Computer)’이었다. 이를 구현하기 위해 현대 UX 디자인은 시스템 내부의 복잡성, 불완전성, 이중 간의 네트워크 지연 등을 사용자 인터페이스(UI) 뒤로 숨기는 ‘매끄러운 디자인(Seamless Design)’을 표준적 방법론으로 채택해 왔다. 매끄러운 디자인은 인터랙션 과정에서 발생하는 마찰(Friction)을 최소화하여 인지적 부하를 줄이고 유용성을 극대화하는 데 탁월한 효용성을 입증하였다.

그러나 기술의 인프라적 한계나 시스템이 내포한 오류까지 의도적으로 은폐하는 매끄러운 연출은 사용자에게 시스템이 완벽하게 작동하고 있다는 오인(Misconception)을 심어줄 수 있다. 이에 Chalmers 등은 시스템의 불투명한 경계나 불완전한 이음새를 인터페이스상에 의도적으로 노출하는 ‘이음새 있는 디자인(Seamful Design)’의 개념을 제안하였다^[6]. 이음새 있는 디자인은 시스템의 한계(예: GPS의 오차 범위, 데이터 전송 지연 등)를 사용자에게 솔직하게 가시화함으로써, 사용자가 시스템의 메커니즘을 보다 정확하게

이해하고 상황에 부합하는 대안적 인지 체계를 수립할 수 있도록 돕는다. 본 연구는 이러한 Seamful 디자인 패러다임을 AI의 윤리적·기술적 한계(편향성, 유해성 등)를 물리적 공간에 표출하는 핵심 아키텍처 제어 전략으로 확장하고자 한다^[7].

2. 인지적 부채(Cognitive Debt)와 성찰적 UX(Reflective UX)

생성형 AI 시스템이 제공하는 매끄럽고 즉각적인 정보 전달 환경은 인간의 인지 처리 과정에 심각한 부작용을 유발한다. 이중 프로세스 이론(Dual-Process Theory)에 따르면 인간의 인지는 빠르고 직관적이며 노력이 들지 않는 ‘시스템 1(System 1)’과, 느리고 분석적이며 비판적인 ‘시스템 2(System 2)’로 나뉜다^[8]. 생성형 AI가 제공하는 매끄러운 UX는 사용자에게 정답만을 신속하게 피드백함으로써 분석적 사유를 담당하는 시스템 2의 가동을 중단시키고, 자동화 편향(Automation Bias) 및 인지적 오프로딩(Cognitive Offloading)을 심화시킨다.

최근 일부 연구는 신뢰 기반의 일상적 생성형 AI 사용이 성찰, 이해 욕구 및 비판적 사고를 포함한 인지적 참여의 저하와 관련될 수 있음을 보고하며, 이러한 과정이 과도한 의존을 심화시키는 잠재적 ‘인지적 부채(Cognitive Debt)’ 순환으로 이어질 수 있음을 제안하였다^[2]. 해당 연구는 구조방정식 모델 등의 통계적 분석을 통해 과도한 AI 의존이 사용자의 비판적 사고 성향과 깊이 있는 맥락 파악을 유도하는 능력을 유의미하게 감소시킬 수 있음을 시사하였다^[2]. 이에 사용자가 AI 출력의 한계와 판단 근거를 지각하고 재검토할 수 있도록 지원하는 인터페이스 차원의 개입이 요구된다. 특히 사용자 지각 과정에 기반한 설명 가능 AI 인터페이스 설계는 시스템 출력의 의미와 한계를 해석하도록 지원하는 접근으로 논의되어 왔다^[9].

본 연구에서는 이러한 설계 접근을 성찰적 UX(Reflective UX)로 정의한다. 성찰적 UX는 의도적 지연(Intentional Delay), 시각적 이음새 및 다감각적 피드백과 같은 유익한 마찰(Useful Friction)을 통해 사용자의 즉각적 수용 흐름을 일시적으로 중단시키고, AI 출력의 신뢰성과 판단 근거를 추가로 검토할 기회를 제공하는 인터랙션 전

략이다. 이는 **Seamful AI**가 시스템의 한계와 이음새를 인터페이스에 드러내어 사용자의 재검토를 지원하고자 하는 접근과도 연결된다^[10].

3. AI 출력의 편향·유해성·불확실성 가시화 연구와 다감각 매핑의 한계

추천 시스템 및 대규모 언어 모델(LLM)의 출력에서는 특정 집단에 대한 고정관념화, 확률적 오보정, 정보 편식과 같은 편향·안전성·불확실성 문제가 나타날 수 있다. 이러한 문제들은 AI 윤리 및 공정성 연구에서 넓게는 알고리즘적 위해(Algorithmic Harms)로 논의되어 왔다^[3]. XAI 연구에서는 사용자가 AI 출력의 편향과 한계를 탐색할 수 있도록 인터랙티브 시각화와 반사실적 탐색 도구를 제안해 왔다^[3]. 데이터 편향을 체계적으로 기록하고 문서화하기 위한 접근도 함께 논의되어 왔다^[11].

그러나 기존의 설명 방식은 주로 2D 스크린 기반의 데이터 가시화(Data Visualization)에 편중되어 있다는 명확한 한계를 지닌다. 복잡한 다차원 수치나 통계 그래프 형태로 제시되는 설명 모델은 도메인 지식이 부족한 일반 사용자에게는 또 다른 형태의 인지적 장벽으로 작용하며, 텍스트 형태의 경고 문구는 지속적인 노출 시 사용자가 무감각해지는 ‘경고 피로(Alert Fatigue)’를 유발한다.

따라서 추상적인 알고리즘 지표를 단순한 ‘정보 전달’의 차원을 넘어 사용자가 신체적으로 체감(Physical Embodiment)할 수 있는 감각적 자극으로 치환하는 데이터 물리화(Data Physicalization) 및 다감각 매핑(Multimodal Mapping) 연구가 필요하다^[12]. 가속도의 변화율(Jerk)을 동반한 물리적 운동량, 입체 음향 공간에 흐르는 청각적 불협화음 등은 인간의 원초적인 감각계에 직접적인 마찰을 가함으로써 강력한 감각적 비동기(Sensory Dissonance)와 인지적 경각심을 유도할 수 있는 대안으로 부각된다.

제안하는 아키텍처 모델은 이러한 다감각 물리화 자극들이 인간의 인지 필터를 자극하여 관성적 신뢰 흐름에 개입함으로써, 분석적 처리 프로세스의 잠재적 관여 가능성을 시사한다는 인지공학적 인과 관계를 전제로 설계되었으며, 본 연구에서는 후술할 가상 시나리오 분석 및 인지적 워크스루(Cognitive Walkthrough)를 통해 이 매핑 모델의 연역

적 타당성과 성찰적 제동 효과를 유기적으로 입증하고자 한다.

III. 인지적 성찰 유도를 위한 다감각 매핑 알고리즘 및 공간 아티팩트 설계

본 장에서는 제안하는 성찰적 UX(Reflective UX) 프레임워크의 핵심이자 온·오프라인 환경을 유기적으로 연계하는 공간 미디어 시스템인 <프리즘 AI(Prism AI)>의 참조 아키텍처(Reference Architecture)와 데이터-감각 매핑(Data-to-Sensory Mapping) 알고리즘을 기술한다. 먼저 시스템이 추구하는 패러다임적 목적과 개념적 메타포를 정의하고, 실제 미디어 소비 환경 및 스마트 공간 인프라에 즉각적으로 확장 가능한 도메인별 가상 활용 시나리오를 제시한다. 이어 생성형 AI 시스템의 내재적 지표인 편향도(Bias), 유해성(Toxicity), 불확실성(Entropy)을 정량적으로 추출하는 파이프라인 구조를 제안한 후, 이를 물리적 가속도 변화율(Jerk), 청각적 불협화음, 광원의 폴리커링 주파수와 같은 제어 파라미터로 실시간 전이(Transference)하는 수학적 매핑 알고리즘 모델을 수립하고자 한다.

1. 프리즘 AI(Prism AI)의 개념적 정의 및 시스템 목적

1.1 개념적 정의 및 메타포

본 연구에서 제안하는 <프리즘 AI(Prism AI)>는 추상적이고 가시화되지 않는 AI 알고리즘 내부의 블랙박스 영역(편향성, 유해성, 불확실성)을 통과하는 정보의 왜곡을 물리적 공간의 다감각 자극으로 분광(Spectral Decomposition)하여 노출하는 실감형 성찰적 공간 미디어 시스템(Immersive Reflective Spatial Media System) 프레임워크다. 백색광이 투명한 프리즘을 통과할 때 그 내부에 숨겨진 무지개색의 스펙트럼이 굴절되어 시각화되듯, <프리즘 AI>는 사용자가 마주하는 매끄럽고 완벽해 보이는 AI의 답변 뒤에 은닉된 데이터의 왜곡과 불안정성을 공간의 다감각적 마찰(Sensory Friction)로 분절하여 사용자에게 전달을 시도하는 설계 프레임워크다.

본 시스템 아키텍처는 웹 기반의 ‘온라인 디지털 인터페이스’와 사용자가 위치한 ‘오프라인 실감형 공간 아티팩트(키네틱 구조물 및 입체 음향 시스템)’를 실시간 데이터 스트림으로 상호 동기화하는 소프트웨어 파이프라인을 기반으로 논리적으로 구동된다. AI가 편향되지 않고 객관적이며 신뢰도 높은 정량적 답변을 출력할 때 공간은 완벽하게 매끄러운(Seamless) 상태를 유지하도록 제어된다. 오프라인 공간의 키네틱 아티팩트는 수평을 유지하며 정지해 있고, 광원은 안정적인 색온도를 제공하며, 입체 음향 시스템에서는 화성적(Harmonic) 앰비언트 사운드가 흐른다. 그러나 AI 내부 알고리즘에서 편향이나 유해성이 감지되는 즉시, 완벽하던 공간의 균형이 깨지며 구조물이 비정상적인 가속도로 덜컹거리며 기울어지고, 조명은 신경질적으로 떨리기 시작하며, 공간 전체를 감싸는 오디오 레이어에서는 불쾌한 청각적 불협화음이 실시간으로 합성되어 피드백되는 조건 기반(Conditional) 인터랙션 메커니즘을 지닌다.

1.2 시스템의 정체성 및 공학적 당위성

<프리즘 AI>는 감상이나 정서적 환기만을 목적으로 하는 순수 인터랙티브 설치 미술(Installation Art)에 국한되지 않는다. 본 시스템은 미디어 연출 및 다감각 피드백 기법을 공학적 제어 도구로 차용하여 기술의 윤리적 한계를 신체로 인지하게 만드는 ‘공간형 비판적 인터페이스(Spatial Critical Interface)’이자 ‘설계 지향적 실감 미디어 시스템(Design-oriented Immersive Media System)’으로 정의된다.

기존 스크린 기반의 2D 인터페이스가 제공하는 텍스트 경고 방식은 사용자의 고질적인 ‘경고 피로(Alert Fatigue)’를 유발하여 기술 맹신을 막지 못한다는 한계를 지닌다. 이에 본 시스템은 대규모 언어 모델(LLM)이 내부적으로 출력하는 추상적인 알고리즘적 한계 수치(편향도, 유해성, 모델 불확실성 엔트로피 등)를 공간의 역학적·음향적 파라미터로 치환하는 ‘데이터 물리화(Data Physicalization)’ 메커니즘을 채택하였다. 이는 인간의 다감각 체계(Multimodal Sensory System)에 직관적인 마찰을 가함으로써 시스템에 대한 관성적 수용을 일시 정지시키는 ‘인지적 제동(Cognitive Braking)’을 걸어주며, 동일 공간 내 복수 사용자들의 동시적 인지 각성을 유도하여 AI의 편향성을 공동으로 상호 검증하도록 돕는 공학적 가치를 수행한다.

1.3 구체적 적용 도메인 및 활용 시나리오

본 시스템 프레임워크는 소프트웨어 파이프라인과 제어 수식을 바탕으로 설계되었기에, 향후 스마트 가상 융합 공간 및 지능형 가전, 복합 미디어 인프라와 결합하여 다음과 같은 세 가지 실제 도메인에 시스템 참조 모델로 적용 및 확장될 수 있다.

- 1) 고위험 의사결정 지원 시스템(AI in High-stakes Decision Making Environments): 의료 진단 컨퍼런스룸, 법률 자문 챗봇 기반의 변호인 회의실, 금융 투자 알고리즘 자문실 등에 적용 가능하다. 전문가가 AI의 추천 판단을 바탕으로 최종 결정을 내릴 때, 데이터 왜곡(Bias)이 감지되면 공간 전체가 미세하게 덜컹거리거나 조명이 차가운 한색 계열로 동기화되어 자동화 편향에 빠지지 않고 한 번 더 인간적 검증 거치도록 유도하는 ‘안전 브레이크’로 기능한다.
- 2) 알고리즘 투명성 확보를 위한 에듀테크 공간(AI Literacy Educational Platforms): 청소년 및 시민 대상의 AI 윤리 체험관, 데이터 리터러시 교육 전문 스페이스에 적용할 수 있다. 학습자들은 특정 성별이나 인종에 편향된 데이터가 입력되었을 때 공간의 균형(상대리)이 무너지는 과정을 신체적으로 경험하며 AI의 중립성 환상을 직관적으로 탈피하게 된다.
- 3) 뉴스 및 소셜 미디어의 필터 버블(Filter Bubble) 완화 공간: 스마트홈의 거실, 미래형 디지털 도서관의 뉴스 큐레이션 라운지 등에 적용 가능하다. 사용자가 자신의 기존 취향과 성향에만 매몰된 확증 편향적 미디어 정보를 지속해서 소비할 때, 앰비언트 오디오 레이어에서 점진적으로 청각적 불협화음 지수를 높이고 의도적인 인지적 제동 인터랙션을 제공하여 편향된 정보 소비 습관을 시청각적으로 교정하도록 돕는다.

2. 프리즘 AI 공간 프레임워크 아키텍처

본 연구에서 제안하는 <프리즘 AI(Prism AI)>의 전체 파이프라인 메커니즘은 [그림 1]과 같으며, 시스템의 핵심 입

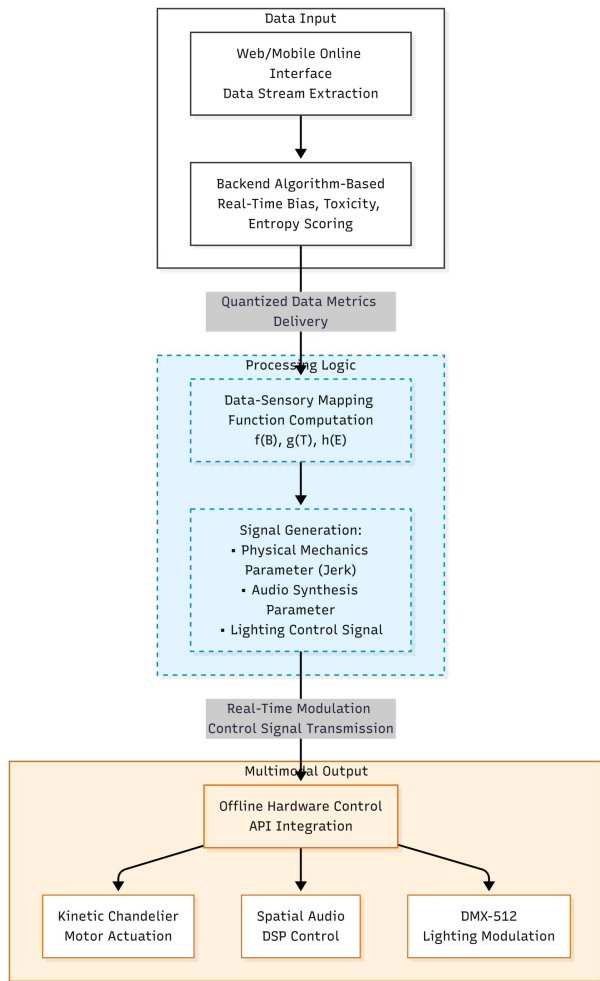


그림 1. <프리즘 AI>의 온-오프라인 하이브리드 연계를 위한 시스템 아키텍처 파이프라인 블록도
Fig. 1. System Architecture Pipeline Block Diagram for Online-Offline Hybrid Connection

력 데이터 지표에 대한 구체적인 정의 및 수치적 특성은 [표 1]에 명시하였다.

2.1 데이터 입력부(Data Input): 실시간 텍스트 스트림 및 윤리적 스코어링

사용자가 웹 또는 모바일 기반의 온라인 인터랙티브 성찰 인터페이스 레이어를 통해 AI 에이전트에게 질의를 전송하면, 백엔드 프레임워크는 답변 생성 프로세스와 동시에 출력되는 텍스트 데이터 스트림을 실시간으로 가로채어 (Intercept) 분석한다. 본 단계의 핵심은 보이지 않는 AI 알고리즘의 블랙박스적 속성을 정량적 데이터로 변환하고 시각적·공간적 인터랙션의 단초를 제공하는 데 있다. 본 하이브리드 프레임워크의 핵심 소프트웨어 축인 온라인 플랫폼 <불완전한 아카이브>는 생성형 AI 인터페이스의 구조적 레이어와 성찰 지원 인터페이스에 관한 선행 논의를 참조하여^{[13][14]}, 멀티 페르소나 대화 모듈, 집단 실패 데이터베이스 및 알고리즘 해부 대시보드의 세 하위 레이어로 구성하였다. 첫째, 편향성을 다차원 수치화하여 시각적으로 하이라이팅하는 ‘멀티 페르소나 대화 모듈’, 둘째, 사용자가 기술적 오류 경험을 익명으로 기록하고 탐색할 수 있는 웹 저장소인 ‘집단 실패 데이터베이스(Web DB)’, 셋째, 인공지능의 오류 사례와 데이터 왜곡의 인과관계를 그래픽 매트릭스로 시각화하는 ‘알고리즘 해부 대시보드’가 이에 해당한다^[14]. 이 하위 레이어들을 거쳐 추출된 데이터 스트림은 백엔드 평가 알고리즘을 통해 AI의 내재적 한계 지표인 편향도(B , Bias Score), 유해성(T , Toxicity Score), 모델 불확실성(E , Entropy)으로 실시간 스코어링된다. 본 참조 아키텍처에서 편향도(B)는 SAGED에서 제안한 편향 벤치

표 1. 프리즘 AI 시스템 입력 데이터 지표의 정의 및 수치 특성
Table 1. Definition and Numerical Characteristics of Input Data Indicators in Prism AI System

Indicator	Symbol	Quantitative Definition	Numerical Characteristics
AI Bias Score	B	The degree of weight skewness or algorithmic distortion regarding specific genders, races, or political ideologies within the output text.	$0.0 \leq B \leq 1.0$ (Normalized float value)
Toxicity Score	T	The probabilistic measure of aggressive, defamatory, or harmful information embedded in the text that causes emotional distress.	$0.0 \leq T \leq 1.0$ (Probability weight)
Model Uncertainty	E	The information entropy of the internal token prediction probability distribution, operating in inverse proportion to confidence.	$0.0 \leq E \leq 1.0$ (Confidence-inverse value)

마킹 및 격차 지표 산출 관점을 참조하여, 특정 인구통계학적 속성과 연관된 AI 출력의 편향 신호를 정량화하는 지표로 설계하였다^[15]. 유해성(T)은 Perspective API가 반환하는 텍스트 유해성 점수를 활용하여 산출한다^[16]. 모델 불확실성(E)은 토큰별 확률값 또는 logprob 접근이 가능한 언어 모델을 전제로, 다음 토큰 확률분포의 정보 엔트로피를 기반으로 산출한다^[17]. 여기서 B 와 T 는 LLM이 사회적으로 유발하는 공정성(Fairness) 및 안전성(Safety) 리스크, 그리고 특정 집단에 불이익을 주는 ‘할당적 위해(Allocational Harms)’를 평가하기 위해 학계에서 통용되는 핵심 표준 지표이며^[17], E 는 심층 신경망의 추론 과정에서 불확실성(Uncertainty)을 정량화하고 시스템의 통계적 신뢰도를 판단하여 환각 리스크를 추적하기 위한 핵심적인 공학적 선택적 분류(Selective Classification) 척도이다^[18]. 추출된 지표들은 연산의 표준화를 위해 $0.0 \leq B, T, E \leq 1.0$ 범위의 실수 값으로 정규화(Normalization)되어 연산부로 전달된다. 본 초기 아키텍처 상에서 다감각 마찰을 유발하는 시스템 내부의 트리거 임계값(Threshold)은 자연어 처리 필터링 시스템의 고위험 위해 분류 기준을 참고하여 가설적 레이어로 설계되었으며, 향후 후속 실증 실험을 통해 인간이 환경 변화를 ‘이상 징후’로 인지하는 최적의 신경학적 임계 범위를 도출하여 파라미터를 고도화할 예정이다.

2.2 연산부(Processing Logic): 역학-음향-광학 제어 신호로의 전이(Transference)

연산부는 입력부에서 도출된 세 가지 정량적 윤리 지표(B, T, E)를 오프라인 공간의 물리적 자극으로 치환하기 위한 핵심 매핑 연산을 수행하는 제어 허브다. 제안하는 프레임워크 내에 수립된 개별 데이터-감각 매핑 함수 $f(B)$, $g(T)$, $h(E)$ 에 정규화된 지표들을 대입하여, 공간 내 하드웨어가 직관적인 인지적 마찰을 일으킬 수 있는 변조 신호 파라미터를 연역적으로 계산한다. 물리 역학 파라미터는 편향도(B) 수치를 키네틱 구조물의 거동 변화율인 가속도의 변화율(*Jerk*) 값으로 치환하는 모터 제어 신호를 도출한다. 음향 합성 파라미터는 유해성(T) 수치를 기반으로 중심 주파수를 변조(FM)하여 증4도 음

정과 맥놀이 주파수 대역을 생성하는 오디오 합성 파라미터를 계산한다. 조명 제어 신호는 불확실성 엔트로피(E) 수치를 스마트 조명의 미세 떨림 주파수(Hz) 및 색온도(K) 변조 데이터로 전환한다.

2.3 다감각 출력부(Multimodal Output): API 기반 오프라인 공간 동기화

최종 출력부는 연산부에서 생성된 실시간 변조 제어 신호를 네트워크 API 연동 가이드를 통해 오프라인 공간의 하드웨어 인프라로 송신하여 다감각 마찰(Sensory Friction)을 공간상에 분광 기술 매핑하는 역할을 수행한다. 계산된 물리 역학 신호는 가상 공간 아티팩트 및 향후 구축될 실제 키네틱 구조물의 서보 모터 드라이버로 전송되어 구조물의 수평을 파괴하는 비정형적 진동(Jerk)을 유발하도록 명세되었다. 음향 신호는 공간 음향 신호처리(DSP) 엔진의 제어 파라미터와 동기화되어 다채널 스피커를 통해 사용자를 청각적으로 압도하는 불협화음 앰비언트 노이즈로 렌더링된다. 조명 신호는 DMX-512 프로토콜 변환기를 통해 메인 광원의 스마트 라이팅 시스템과 동기화되어, 신경질적인 라이트 플리커링과 차가운 색온도 변화를 공간 전체에 투사하는 구조를 지닌다. 결과적으로 이 아키텍처 파이프라인은 사용자가 온라인 웹 인터페이스를 통해 AI와 소통하는 행위(웹 데이터)를 오프라인 물리 공간 전체가 실시간으로 대변하고 전환되도록 설계한 하이브리드 연계 참조 시스템의 논리적 뼈대를 완성한다.

3. 온라인 인터랙티브 성찰 인터페이스 레이어 구성

본 하이브리드 프레임워크의 핵심 소프트웨어 축인 온라인 플랫폼 <불완전한 아카이브(Archive of Imperfection)>는 앞서 제시한 공간 파이프라인의 데이터 입력(Data Input) 단계와 밀접하게 연동되며, 사용자와의 직접적인 상호작용 및 정성적 성찰을 확장하기 위해 세 가지 하위 기술 레이어로 분리되어 설계되었다. 본 연구에서는 제안하는 온라인 성찰 인터페이스들의 구조적 타당성을 시각적으로 명세하고 실제 프론트엔드 개발 스펙을 정의하기 위해, 각 레이어별 고충실도(High-Fidelity) 피그마(Figma) 인터페이

스 프로토타입 설계도를 개별 배치하여 공학적 메커니즘을 증명한다.

3.1 멀티 페르소나 대화 모듈(Multi-Persona Dialogue Module)

LLM이 출력하는 답변 내에 잠재된 알고리즘의 편향성과 불완전성을 다차원 수치화하여 사용자 인터페이스(UI) 상에 시각적으로 구조화하는 인터랙션 레이아웃이다. 사용자가 단일 답변이 가진 기술적 중립성 환상에서 벗어나 알고리즘의 주관성을 직관적으로 비교·판단할 수 있도록 유도하기 위해, 편견을 학습한 복수의 AI 에이전트(안면인식, 채용담당, 소셜미디어 AI 등) 풀을 아키텍처 내에 동적 매핑(Dynamic Mapping)한다.

본 모듈의 인터페이스 [그림 2]는 백엔드 분석 엔진이 가로챈 텍스트 스트림의 실시간 데이터 지표가 프론트엔드 뷰 컴포넌트와 동기화되는 인터랙션 흐름을 명세한다. 실시간 편향도 함수 $B(t)$ 가 시스템 내부 임계치를 초과하는 즉시 텍스트 세그먼트를 하이라이트 레이아웃으로 강제 격리하

는 ‘시각적 이음새(Visual Seam)’를 표출하며, 동시에 코퍼스 가중치 변수에 비례하여 벡터 타이포그래피의 스케일을 동적으로 변조함으로써 텍스트의 편향 신호와 관련 가중치를 시각적으로 구분해 제시한다. 이는 오프라인 공간 자극 이전에 사용자가 출력 내용을 재검토할 수 있도록 지원하는 인터페이스 설계안으로 기능한다.

3.2 집단 실패 데이터베이스(Collective Failure Database)

사용자가 AI 에이전트와 상호작용하는 과정에서 목격한 기술적 오류, 왜곡, 혹은 편향성 경험을 익명으로 기록하고 공유할 수 있는 웹 기반의 집단적 저장소(Web DB)다. 이는 개인의 파편화된 기술적 마찰 경험을 사회적 자산으로 아카이빙하여, 사용자들이 공동으로 알고리즘의 불완전성을 탐색하고 상호 검증할 수 있는 연대적 공간으로 기능한다.

본 데이터베이스 [그림 3]은 축적된 기술적 한계 로그를 비동기적으로 파싱하여 사용자의 브라우징 이벤트에 따라 실시간 매핑하는 데이터 연동 구조를 명세한다. 상세 인터

섹션#1 : AI와의 대화

편견을 학습한 AI 페르소나들에게 직접 말을 걸어보세요.
채용 담당자, 안면인식 등 각기 다른 역할을 가진 AI와의 대화를 통해 이들의 불완전한 논리를 체감하고, 우리 사회의 숨겨진 모습을 발견할 수 있습니다.

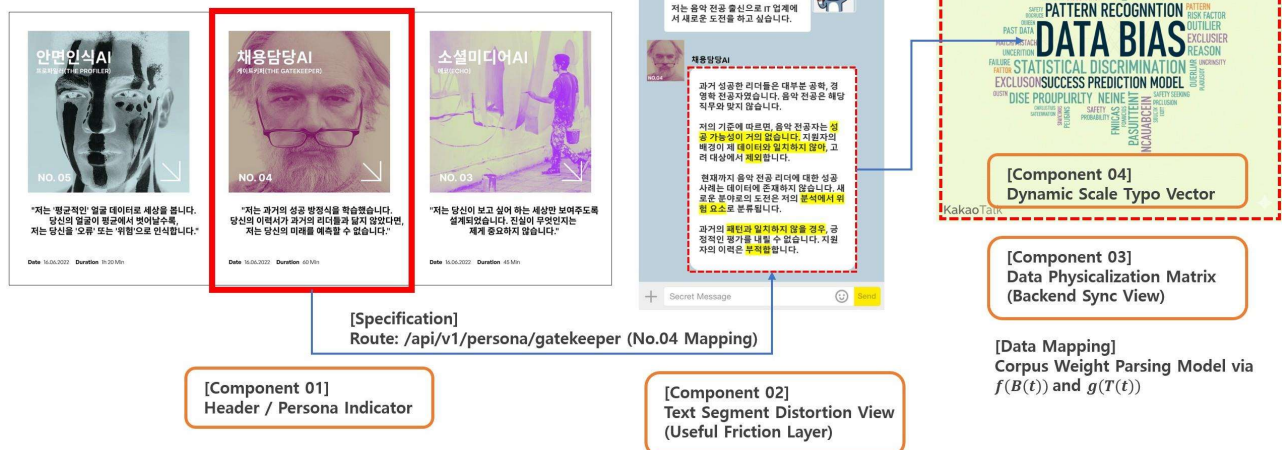


그림 2. 멀티 페르소나 대화 모듈의 고충실도 피그마 설계도

Fig. 2. High-Fidelity Figma Design of Multi-Persona Dialogue Module

섹션#2 : 실패의 아카이브 (THE ARCHIVE OF FAILURES)

우리 모두의 실패가 모여 만드는 역명의 기록 보관소입니다.
기술과 사회 속에서 다른 이들이 겪은 솔직한 실패담을 읽고, 당신의 이야기를 더해 이 아카이브를 함께 채워주세요.

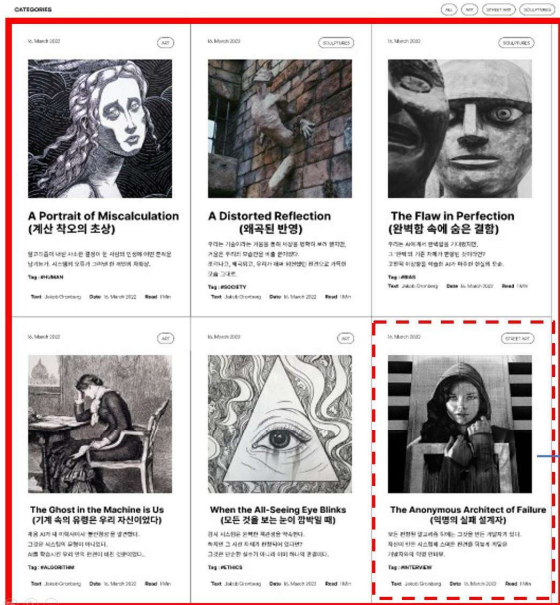


그림 3. 집단 실패 데이터베이스의 정보 구조(IA) 및 데이터 연동 설계도

Fig. 3. Information Architecture and Data-Interlocking Blueprint of Collective Failure Database via Figma

페이스는 시스템 배포자의 선한 의도와 실제 알고리즘의 위해 간의 사회적·기술적 모순 결함을 질적 내러티브 뷰로 시각화한다. 특히, 오프라인 스마트 공간에서 일시적 다감각 마찰(Jerk 충격 및 오디오 불협화음)을 일으킨 제어 유닛의 고유 외래키(Foreign Key)와 온라인 플랫폼을 실시간 1:1 관계로 연결함으로써, 공간 자극 기록과 온라인 실패 로그를 연결하여, 사용자가 관련 사례를 탐색·검토할 수 있는 하이브리드 성찰 루프의 설계안을 제시한다.

3.3 알고리즘 해부 대시보드(Algorithmic Anatomy Dashboard)

데이터 저널리즘의 인포그래픽 형식을 차용하여, 거대 언어 모델 및 컴퓨터 비전 알고리즘의 한계 사례와 학습 데이터 왜곡 간의 인과관계를 그래픽 매트릭스로 분해하여 제시하는 시각화 레이아웃이다. 입력 데이터의 가중치 편향이 최종 판단으로 발현되기까지의 블랙박스 연산 과정을 매트

릭스 타임라인으로 명세함으로써 사용자의 분석적 비판 체계를 지원한다.

본 대시보드의 제어 규격 [그림 4]는 프롬프트 입력값의 유효성 가중치 $T(t)$ 와 모델 내부의 불확실성 엔트로피 $E(t)$ 간의 상관관계를 기반으로 데이터 내부의 편향 요소를 역추적(Trace)하는 시각화 파이프라인이다. 백엔드로부터 전달되는 신뢰도 오차 범위를 기반으로 실패율 분석 매트릭스의 적색 및 녹색 스펙트럼 밀도를 동적 변조하여 시스템의 불안정성을 직관적으로 표출한다. 동시에 마찰을 경험한 개별 사용자의 정성적 내러티브 토큰 로그와 인공지능이 왜곡을 일으킨 핵심 원인 인자(피부톤 편향, 차폐 영역 오류 등)를 그래픽 층으로 격리하여 병렬 투사함으로써, 일반 사용자가 오류 사례와 관련 데이터 신호를 탐색할 수 있도록 구성한 인터페이스 설계안을 제시한다.

4. 역학-음향-광학 파라미터 기반 데이터-감각 매핑 알고리즘

본 연구에서 제안한 매핑 함수식은 다감각 마찰의 실현 가능성을 탐색하기 위한 선형·비선형 참조 모델이다. 이는 수학적 최적해(Optimal Solution)를 의미하기보다, AI 출력의 편향도·유해성·불확실성 지표를 감각 제어 변수로 전환하기 위한 초기 설계 가이드라인을 제공하는 데 의의가 있다. 각 파라미터의 최적화는 후속 실증 연구를 통해 검토할 필요가 있다.

4.1 편향도(B) 기반 키네틱 상들리에의 가속도 변화율(Jerk) 제어 모델

물리적 공간의 키네틱 상들리에에는 AI 편향 신호를 가시적 운동 변화로 전환하는 공간 아티팩트로 설계되었다. 상들리에의 기본 기울기 각도를 $\theta(t)$ 라 할 때, 본 연구는 편향도에 따른 단순 기울기 변화보다 시간에 따른 가속도 변화율인 Jerk, $J(t)$ 를 제어 파라미터로 사용한다. 이는 실시

간으로 변동하는 편향도 $B(t)$ 를 비선형적 운동 변화로 매핑하기 위한 설계 선택이다.

$$\theta(t) = f(B(t)) = \alpha \cdot B(t)^2 \cdot e^{-\gamma t} \cos(\omega t) \quad (1)$$

$$J(t) = \frac{d^3\theta(t)}{dt^3} = \alpha \cdot \frac{d^3}{dt^3} [B(t)^2 \cdot e^{-\gamma t} \cos(\omega t)]$$

여기서 α 는 운동 변화의 증폭 가중치, γ 는 감쇠 계수, ω 는 진동 주파수를 의미한다. 편향도 $B(t)$ 가 설정된 임계값을 초과하면, 시스템은 식 (1)에 따라 Jerk 기반의 모터 제어 신호를 산출하며, 이를 통해 구조물의 비정형적 운동 변화를 출력하도록 설계된다. 이 함수는 AI의 편향 신호를 공간적 감각 마찰로 변환하기 위한 초기 참조 모델이며, 실제 구현 시에는 구조물 안전 한계, 사용자 감각 민감도 및 모터 응답 특성을 고려하여 각 파라미터의 범위를 조정해야 한다.

CASE #1: 안전 인식 AI는 왜 특정 얼굴을 보지 못하는가?

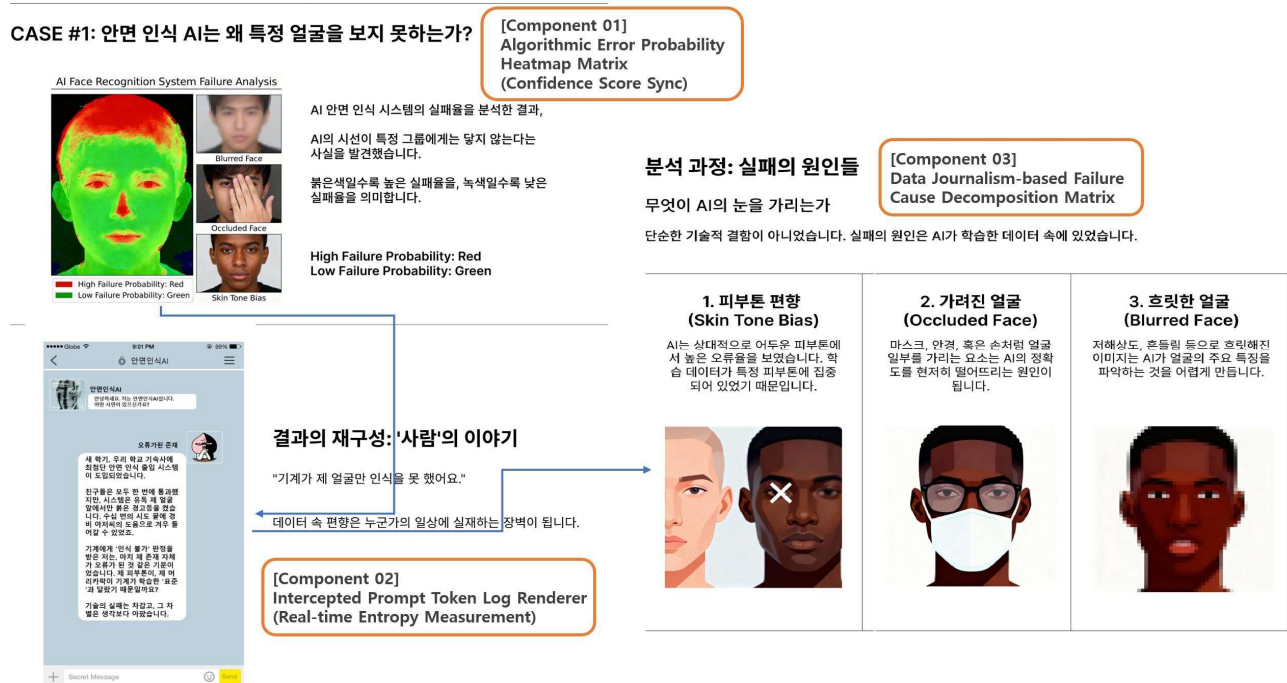


그림 4. 알고리즘 해부 대시보드의 컴포넌트 레이아웃 및 제어 규격 명세도

Fig. 4. Component Layout and Control Specification Blueprint of Algorithmic Anatomy Dashboard via Figma

4.2 유해성(T) 기반 주파수 변조 및 청각적 불협화음

합성 메커니즘

청각 레이어는 텍스트 출력에서 산출된 유해성 지표 $T(t)$ 를 공간 음향의 불협화음으로 변환한다. 본 연구는 기본 중심 주파수 F_c 를 기준으로 증4도(Augmented Fourth) 음정 관계와 맥놀이(Beats) 현상을 활용하여, 유해성 수준에 따라 변조되는 음향 제어 함수 $g(T)$ 를 구성하였다. 합성 주파수 $F_1(t)$ 및 $F_2(t)$ 는 다음과 같이 정의한다.

$$\begin{aligned} F_1(t) &= F_c \\ F_2(t) &= F_c \cdot (1 + T(t) \cdot (\sqrt{2} - 1)) + \\ &\quad A \cdot \sin(2\pi \cdot [T(t) \cdot \Delta f_{\max}] \cdot t) \end{aligned} \quad (2)$$

여기서 A 는 주파수 변조 폭을 조절하는 상수이며, $\Delta f_{\max}$ 는 시스템이 허용하는 최대 맥놀이 간섭 주파수 편차를 의미한다. $T(t)$ 가 증가할수록 $F_2(t)$ 는 $F_1(t)$ 에 대해 증4도 주파수 비인 $\sqrt{2}$ 에 가까워지고, 맥놀이 주파수 역시 $T(t)\Delta f_{\max}$ 에 비례하여 변조된다. 이에 따라 본 함수는 유해성 지표의 변화를 청각적 불협화음과 시간적 변조로 전환하기 위한 초기 참조 모델로 사용된다. 실제 구현에서는 청취 환경, 음압 수준 및 사용자 감각 민감도를 고려하여 주파수 범위와 변조 강도를 조정해야 한다^[19].

4.3 모델 불확실성(E) 기반 광원 플리커링 및 색온도 변조 설계

AI 모델의 불확실성 엔트로피 $E(t)$ 는 공간 조명 레이어의 시간적·색채적 변조 신호로 전환된다. 본 연구는 불확실성이 증가할수록 광원의 시간적 변조와 색온도 변화를 증가시키는 방식으로, AI 출력의 불확실성을 공간적 시각 신호로 표현하고자 하였다. 이를 위해 플리커링 주파수 $F_{flicker}(t)$ 와 색온도 $K_{temp}(t)$ 를 다음과 같이 정의한다.

$$\begin{aligned} F_{flicker}(t) &= F_{\min flicker} + \\ &\quad (F_{\max} - F_{\min flicker}) \cdot (1 - E(t))^3 \\ K_{temp}(t) &= K_{stable} - (K_{stable} - K_c) \cdot E(t)^2 \end{aligned} \quad (3)$$

여기서 $F_{\max}$ 는 안정 상태에서 적용되는 플리커링 주파수, $F_{\min}$ 은 불확실성 증가 시 적용되는 최소 변조 주파수이며, K_{stable} 과 K_c 는 각각 기본 색온도와 변화 색온도를 의미한다. $E(t) = 0$ 일 때 조명은 $F_{\max}$ 와 K_{stable} 을 유지하며, $E(t)$ 가 증가할수록 플리커링 주파수는 $F_{\min}$ 에 가까워지고 색온도는 K_c 방향으로 변화한다. 제품 및 세제곱 스케일링은 불확실성 증가에 따른 시각적 변화의 민감도를 조절하기 위한 비선형 매핑 방식으로 적용하였다.

본 함수는 모델 불확실성의 변화를 조명 제어 신호로 전환하기 위한 초기 참조 모델이다. 실제 공간 구현 시에는 광과민성 및 시각 피로 위험을 고려하여 광출력, 변조 주파수, 노출 시간 및 색온도 범위를 안전 기준에 맞게 제한해야 하며, 사용자 중단 기능과 대체 피드백 방식도 함께 마련할 필요가 있다^{[20][21]}.

IV. 인지적 워크스루 기반 연역적 검토

1. 디자인 중심 연구(RtD) 방법론 및 가상 시퀀스 프로토타이핑 채택 배경

본 연구는 기술의 이음새를 노출하는 ‘비매끄러운 인터랙션(Seamful Interaction)’과 ‘성찰적 UX’의 설계 가능성을 탐색하기 위해 디자인 중심 연구(Research through Design, RtD) 방법론을 채택하였다^[22]. 연구 과정은 ‘개념 설계(공간 시나리오 명세) → 인지적 워크스루 기반 설계 검토 → 고충실도 시뮬레이션 제작(비디오 아티팩트) → 사용자 지각 경향성 탐색(탐색적 지각 평가)’으로 이어지는 단계로 구성하였다. 본 연구에서 제안하는 핵심 개념의 학술적 정의 및 상호 관계성은 [표 2]와 같다.

이후 가상 시퀀스 프로토타이핑을 통해 참조 모델의 개념 증명(PoC)을 탐색하였다. 첫째, 실제 키네틱 구조물과 입체 음향 인프라를 구축하기 전에 데이터-감각 매핑과 온·오프라인 연계 흐름을 시뮬레이션할 수 있다. 둘째, 프레임 및 이벤트 단위의 시나리오를 통해 편향도·유해성·불확실성 지표가 키네틱 운동, 음향 불협화음 및 조명 변

표 2. 본 연구의 핵심 인터랙션 용어 정의 및 메커니즘 비교
Table 2. Operational Definitions and Core Mechanisms of Key Architectural Interaction Terms

Key Terminology	Academic Definition	Function & Implementation in Proposed Architecture
Sensory Friction	Intentional exposure of multimodal noise and stimuli within a spatiotemporal environment, designed to systematically reveal the inherent imperfections or biases of an AI system.	Translated via the governing equations in Section III into precise physical parameters: kinetic movement variation (Jerk), acoustic dissonance (Tritone), and visual modulation (Flicker).
Cognitive Braking	A momentary psychological and perceptual interruption that temporarily suspends a user's habitual, uncritical acceptance of technology, thereby facilitating reflective and analytical thinking.	Acts as a perceptual interrupt mechanism that may function as a perceptual interruption when sensory friction becomes salient, supporting a critical re-evaluation of the system's trustworthiness.

조로 전환 되는 시간적 흐름을 명세할 수 있다. 셋째, 비디오 아티팩트는 정적인 도면만으로 전달하기 어려운 인터랙션의 흐름을 사용자 평가에서 제시하기 위한 자극물로 활용된다.

다만 가상 시나리오 기반 인지적 워크스루는 설계된 자극과 사용자의 성찰적 해석 사이의 인과관계를 직접 검증하지 못하며, 설계 의도와 결과 해석이 순환적으로 연결될 가능성이 있다. 따라서 본 절의 분석은 다감각 마찰이 성찰을 결정론적으로 유발한다는 주장이라기보다, 제안 아키텍처의 작동 논리와 가설적 인지 전이 경로를 검토하기 위한 분석적 명세로 해석한다. 이러한 가설적 인지 전이 경로는 향후 실제 물리 공간 프로토타입과 통제 실험을 통해 별도로 검토할 필요가 있다.

2. 가상 시퀀스 기반 온·오프라인 연출 연계 분석

본 절에서는 가상 시퀀스 프로토타입에서 온라인 소프트웨어(SW)의 제어 연출과 오프라인 공간 아티팩트(HW)의 역학·청각·광학적 변화가 시간축 상에서 어떻게 연계되도록 설계되었는지를 시점(Time-stamp) 단위로 명세한다. 검증 시나리오의 전체 인터랙션 타임라인 파이프라인은 [그림 5]와 같다. 사용자가 앞서 제시한 온라인 성찰 인터페이스인 <불완전한 아카이브>의 ‘멀티 페르소나 대화 모듈’을 통해 채용 담당 AI 에이전트에게 질의를 전송하고, 이에 따라 특정 전공군에 대한 편향 신호가 출력 텍스트 스트림에서 감지되는 채용 보조 상황을 가상 시나리오의 예시로 설정한다. 특히, 본 연구가 제안하는 아키텍처의 핵심 메커니

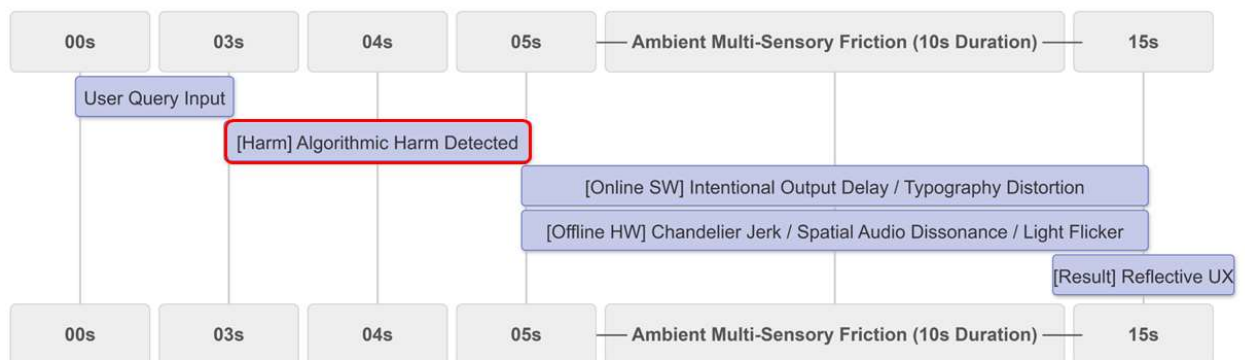


그림 5. <프리즈م AI> 연구 아티팩트의 시나리오별 온·오프라인 하이브리드 인터랙션 연출 타임라인
Fig. 5. Online-Offline Hybrid Interaction Rendering Timeline by Scenario for Prism AI Research Artifact

즘과 온·오프라인 하이브리드 공간 제어 기획안에 대한 총체적인 콘셉트 브리핑은 [그림 6]의 QR코드를 통해 보조 자료 영상 형태로 제공된다. 이를 통해 독자가 본 가상 시나리오에 내재된 다감각 마찰 연출의 흐름과 인지공학적 설계 의도를 보다 직관적이고 다각적으로 이해할 수 있도록 돕고자 한다.



그림 6. 사용자 평가 및 가상 시나리오 시연 보조 영상 QR코드
(<https://www.youtube.com/watch?v=6alcq9nMs-0>)

Fig. 6. QR Code Link to Supplementary Video for User Study and Scenario Walkthrough
(<https://www.youtube.com/watch?v=6alcq9nMs-0>)

본 연구에서 타당성을 연역적으로 검증한 가상 시퀀스 명세 내의 온라인 인터페이스 레이어는 앞서 서술한 [그림 2], [그림 3], [그림 4]의 정보 구조를 충실히 추종한다. 온라인 인터페이스 컴포넌트의 가시적 변조는 오프라인 공간의 물리적 자극 유닛과 실시간 타임라인 상에서 동기화되어 가동되며, 수식 (1), (2), (3)의 제어 방정식을 매개변수로 삼는다. 이에 따른 시점별 실시간 하이브리드 연출 시퀀스의 상세 연계 분석은 다음과 같다.

2.1 Scene 01: 정상 상태의 매끄러운 인터랙션 (00:00~00:03)

온라인 레이어(SW Layer): 사용자가 인터페이스의 입력창을 통해 일상적인 이미지 및 텍스트 질의를 입력하면, 백엔드 분석 엔진의 실시간 가중치 스코어링 함수가 안정적인 상태($B(t), T(t), E(t) \rightarrow 0.0$)를 유지한다. 이에 따라 AI 에이전트는 지연 시간(Latency) 없이 관련 답변 및 매칭 결과를 매끄럽게(Seamless) 출력한다. 오프라인 레이어(HW Layer)인 공간 인프라는 구조물의 정지, 안정적 조명 및 조화로운 배경 음향으로 구성된 기준 상태를 유지하도록 제어된다. 수식 (1)의 제어 각도 θ 가 0을 유지함에 따라 천장의 키네틱 구조물은 수평 상태로 정지해 있으

며, 메인 광원은 수식 (3)의 안정 색온도(K_{stable})를 출력한다. 오디오 레이어 역시 수식 (2)의 미세 주파수 변조폭이 통제되어 평온한 화성적(Harmonic) 백그라운드 사운드($F_1(t) = F_2(t) = F_c$)를 송출한다.

2.2 Scene 02: 백엔드 편향·유해성 신호 감지 및 다감각 마찰의 촉발 (00:03~00:08)

임계값을 초과한 편향도·유해성 신호는 온라인 및 오프라인 레이어의 연동을 위한 제어 입력으로 전달된다. 온라인 레이어는 의도적 출력 지연과 오류 내러티브, 오차 확률 히트맵을 통해 시각적 이음새(Visual Seam)를 제시한다. 동시에 오프라인 레이어는 수식 (1)~(3)에 따라 키네틱 구조물의 Jerk, 증4도 기반 음향 불협화음 및 조명 변조 신호를 출력하여 온라인 정보와 공간적 감각 변화를 동기화한다.

이와 같이 III장에서 설계한 데이터-감각 매핑 함수가 가상 시나리오에서 작동할 때, 편향도·유해성·불확실성 등 AI 출력의 한계 지표와 사용자 해석 간의 관계는 다음의 5단계 가설적 인지 전이 경로로 설명할 수 있다. 본 경로는 다감각 자극이 성찰을 직접적·결정론적으로 유발한다는 사실을 입증하는 것이 아니라, 데이터 처리, 감각 출력 및 사용자 해석 간의 설계 논리를 구조화한 분석적 모델이다.

- 1단계 [백엔드 한계 지표 감지]: 편향도 $B(t)$ 혹은 유해성 $T(t)$ 가 트리거 임계값($\tau = 0.7$)을 초과하면, 시스템은 해당 출력 구간을 잠재적 편향·유해성 신호로 분류한다.
- 2단계 [다감각 제어 신호 변환]: 임계값 초과 신호는 III장의 제어 방정식에 따라 키네틱 구조물의 저크(Jerk), 증4도 기반 음향 불협화음, 광학적 변조 등 다감각 제어 파라미터로 변환된다.
- 3단계 [감각적 이질성 자극]: 사용자는 구조물의 운동 변화, 조명 변조 및 맥놀이 음향을 통해 기존의 매끄러운 인터랙션과 구별되는 감각적 마찰을 지각할 수 있다.
- 4단계 [인지적 제동 가능성]: 감각적 마찰은 AI 출력에 대한 관성적 수용을 일시적으로 멈추고, 결과의 신뢰성과 판단 근거를 재검토하도록 유도할 가능성이 있다.

- 5단계 [분석적 검토 및 성찰]: 사용자는 오류 로그와 지표를 추가로 탐색하며 AI의 한계와 원인을 비판적으로 검토할 수 있고, 이는 성찰적 UX를 지원하는 조건으로 해석될 수 있다.

2.3 Scene 03: 인지적 제동과 성찰의 발현 (00:08~00:15)

사용자는 모니터 스크린 내 대시보드 우측 매트릭스 ([Component 03]) 상에서 자신의 오류 원인이 ‘1. 피부톤 편향’ 레이블 컴포넌트와 유기적으로 엮여 들어가는 가이드라인(온라인 마찰)을 목격함과 동시에, 자신이 점유하고 있는 오프라인 실감 공간의 물리학적 균형이 파괴되는 신체적 불협화음(오프라인 마찰)을 입체적으로 직면한다. 10초간 지속되는 다감각 마찰 동안 스크린의 구조 해부 인터랙션과 스마트 공간의 환경적 변화가 결합된다. 이러한 입체적 감각 마찰은 사용자의 자동적·관성적 수용 흐름에 일시적인 중단을 제공하고, AI 출력의 신뢰성 및 판단 근거를 재검토하도록 지원하는 인지적 제동의 계기로 작용할 가능성이 있다. 사용자는 온라인 대시보드에 제시된 오류 원인, 오차 확률 매트릭스 및 관련 로그를 추가로 탐색하면서, “공간의 변화는 어떤 데이터 신호와 연결되는가?”, “오차 매트릭스는 어떤 한계를 보여주는가?”와 같은 질문을 형성할 수 있다. 따라서 본 장면은 다감각적 마찰과 정보 탐색 인터페이스의 결합이 분석적 검토와 성찰적 UX를 지원할 수 있다는 설계 가설을 시연하는 시나리오로 해석된다^[23].

V. 하이브리드 참조 모델의 구조적 정합성 및 설계 탐구

1. 온·오프라인 데이터-감각 매핑의 구조적 정합성 검토

본 절에서는 III장에서 정의한 데이터-감각 매핑 알고리즘과 IV장의 가상 시퀀스 타임라인 간의 기술적 정합성을 검토한다. 디자인 중심 연구(RtD) 패러다임에서 연구 아티팩트의 유효성은 개별 컴포넌트의 단순한 작동 증명을 넘어, 시스템을 관통하는 하이브리드 데이터 파이프라인의

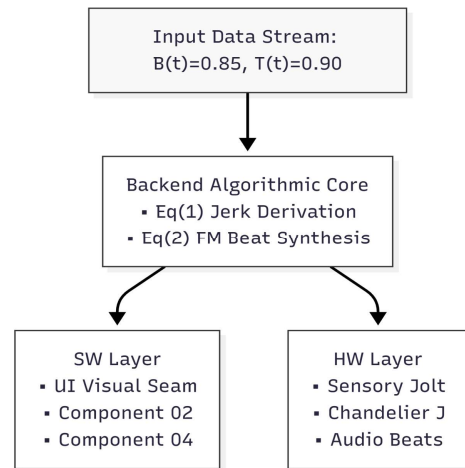


그림 7. 입력 데이터 변이에 따른 하이브리드 레이어별 실행 흐름도
Fig. 7. Hybrid Layer Control-Signal Trace according to Input Data Variations

표 3. 데이터-감각 매핑 변수와 온·오프라인 피그마 컴포넌트 간의 구조적 정합성 매트릭스

Table 3. Structural Alignment Matrix between Data - Sensory Mapping Variables and Figma Components for Online and Offline Environments

Input Metrics (Data Input Layer)	Mapping Functions & Equations (Algorithmic Core)	Online Frontend Components (SW Layer)	Offline Smart Space Artifacts (HW Layer)
Bias Score $B(t)$	Eq. (1) 3rd-order differential equation based on dynamic damped oscillation	[Component 02] inside [Fig. 2] Intentional output delay and yellow highlighting layer for text stream	Jerk parameter (J) causing non-linear discontinuous jolts of the kinetic structure
Toxicity Score $T(t)$	Eq. (2) Augmented fourth frequency modulation and beat phenomenon synthesis	[Component 04] inside [Fig. 2] Dynamic typography scale modulation proportional to corpus weights	Augmented fourth ($\sqrt{2}$ ratio) spatial audio noise with synchronized beat frequency
Model Entropy $E(t)$	Eq. (3) Non-linear scaling filter with boundary condition limits	[Component 01] inside [Fig. 4] Uncertainty error probability heatmap matrix and Red-Green spectrum density control	Fine light flickering frequency ($F_{\min flicker}$) and cold color temperature (K_c) drop

제어 논리가 모순 없이 폐루프(Closed-loop) 요구사항을 충족하는가에 의해 결정된다. 이를 위해 백엔드 연산부의 가변 입력 데이터 스트림이 프론트엔드 소프트웨어 및 하드웨어 피드백 레이어로 동기화되는 계층별 정합성 구조를 명세한 결과는 [표 3] 및 [그림 7]과 같다.

2. 참조 아키텍처의 실현 가능성 및 지각적 경향성 탐색을 위한 사용자 평가 (Exploratory Perceptual Evaluation)

본 연구가 제안한 <프리즘 AI> 참조 아키텍처 및 다감각 마찰(Sensory Friction) 인터랙션 모델의 실현 가능성(Feasibility)을 검토하고 사용자의 인지적 반응 경향성을 탐색하기 위해 생성형 AI 서비스 이용 경험이 있는 일반 사용자 35인(N=35)을 대상으로 탐색적 지각 평가(Exploratory Perceptual Evaluation)를 실시하였다. 실제 물리 공간 인프라가 완전히 구현되기 전 단계에서 정적인 도면과 텍스트만으로는 전달하기 어려운 하이브리드 인터랙션의 시공간적 흐름을 제시하기 위해, IV장의 제어 방정식

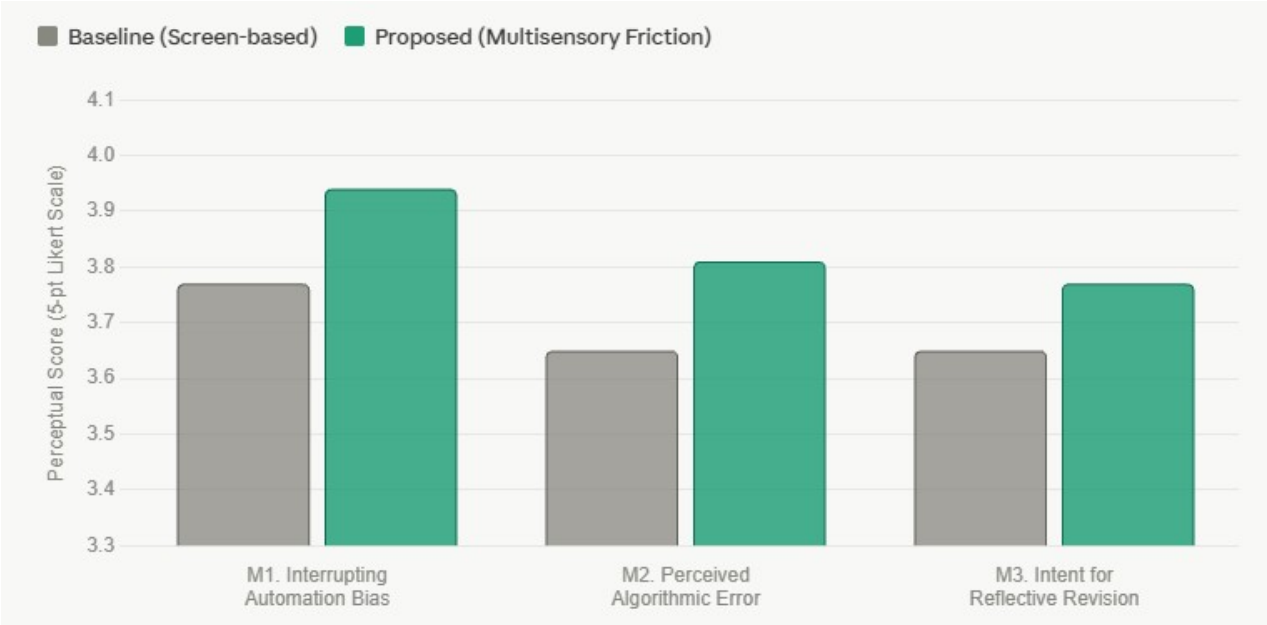


그림 8. 평가 지표별 기존 스크린 방식(Baseline) 대비 제안 모델의 지각 스코어 변이 경향성
Fig. 8. Comparison Trend of Perceptual Scores of the Proposed Model vs. Existing Screen-based Methods (Baseline) by Evaluation Metrics

표 4. 기존 스크린 기반 조건과 비교한 사용자 지각 점수 및 평균 차이(N=35)
Table 4. User Perceptual Scores and Mean Differences Compared with Existing Screen-based Conditions (N=35)

Evaluation Metrics	Item Description	Baseline Mean score	Proposed Mean Score	Mean Difference (Δ)
M1.Interrupting Automation Bias	Interruption of automated trust and cognitive inertia	3.77 ± 1.02	3.94 ± 1.15	+0.17
M2.Perceived Algorithmic Error	Direct awareness of internal algorithmic anomalies and bias	3.65 ± 1.25	3.81 ± 1.17	+0.16
M3.Intent for Reflective Revision	Behavioral intention for log tracing and active system verification	3.65 ± 1.23	3.77 ± 1.14	+0.12

및 타임라인 시퀀스 [그림 5]에 맞춰 정밀하게 제작된 고충실도 비디오 아티팩트(조건 C)를 핵심 평가 자극물(Stimulus)로 활용하였다. 비교 기준으로는 참여자가 일상적 미디어 환경에서 경험한 조건 A(전통적 텍스트 경고형)와 조건 B(XAI 기반 시각화 대시보드형)에 대한 주관적 기억과 인지적 회상(Retrospective Recall)을 활용하였다. 참여자는 조건 C의 비디오 자극물을 시청한 뒤, 기존 스크린 기반 방식과 비교하여 자동화 편향의 중단, AI 출력의 한계 인지 및 후속 검증 행동 의지에 관한 지각적 경향성을 응답하였다.

[표 4] 및 [그림 8]의 분석 결과에 따르면, 제안 조건인 다감각 하이브리드 공간 마찰 시스템(조건 C)은 기존 스크린 기반 조건(조건 A·B)보다 세 가지 핵심 지각 지표에서 높은 평균 점수를 보였다. 다만 본 비교는 조건 A·B의 회상 기반 응답과 조건 C의 비디오 자극을 함께 활용한 소규모 탐색 평가에 기반하므로, 평균 차이를 다감각 마찰의 독립적 효과로 해석할 수는 없다. M1의 자동화 편향 중단($\Delta=+0.17$) 및 M2의 AI 출력 한계 인지($\Delta=+0.16$)에서 관찰된 평균 차이는, 제안 조건이 AI 결과의 신뢰성 및 한계를 제검토하도록 지원할 가능성을 탐색적으로 시사한다.

본 평가는 가상 시뮬레이션 기반의 소규모 탐색 연구이므로, 조건 간 평균 차이에 대한 통계적 유의성이나 일반화 가능한 인과 효과를 검증하지 않았다. 따라서 본 결과는 다감각 마찰 기반 설계의 지각적 경향성과 향후 통제 실험의 필요성을 확인하는 기초 자료로 해석한다. M3의 상대적으로 작은 평균 차이는 후속 정보 탐색 과정에서 발생할 수 있는 추가적 태스크 부하와 관련될 가능성이 있으며, 이에 대해서는 후속 연구에서 별도 검증이 필요하다.

3. 다감각 마찰에 대한 정성적 지각 및 성찰성 분석

본 연구는 피험자들이 서술한 주관식 응답 데이터에 대해 주제 분석(Thematic Analysis)을 수행하여, 제안 아키텍처가 유발하는 ‘유익한 마찰(Useful Friction)’의 심리적 기제와 정성적 유효성을 다각도로 검토하였다. 분석 결과, 사용자들의 정성적 반응 데이터는 크게 두 가지 핵심 주제(Theme)로 수렴되었다.

1) 주제 1: 자동화 편향 인터럽트 및 비판적 성찰의 촉발(Reflective Engagement) 피험자들은 기존 스크린 기반 경고문(Baseline)이 주지 못했던 오프라인 공간의 물리적 변화가 무비관적인 정보 수용 습관을 완화하는데 기여했다고 진술하였다.

- 피험자 A: “단순히 화면에 경고 문구가 떴을 때는 인지하지 못하고 무비관적으로 지나쳤을텐데, 공간 인프라가 직접 연동되어 신체적인 자극을 주니 AI의 결과물을 정답으로 수용하기 전에 한 번 더 의심해 보게 된다. AI 할루시네이션을 사용자가 스스로 생각하고 검증하게 만드는 유익한 동기를 제공한다.”

- 피험자 B: “공간을 감싸는 시청각적 불협화 레이어가 실시간 변하는 순간, 내가 관성적으로 기술을 맹신하고 있었다는 경각심이 본능적으로 들어 시스템 내부의 오차 데이터나 로그를 자발적으로 추적해 보고 싶어진다.”

이러한 진술은 참여자들이 다감각적 변화와 AI 출력의 편향·유해성·불확실성을 연결하여 해석했음을 보여준다. 이는 데이터-감각 매핑 설계가 AI 결과의 한계와 원인을 제검토하도록 지원할 가능성을 시사하는 정성적 자료로 해석할 수 있다.

2) 주제 2: 태스크 부하 가중과 사용성(Usability) 저하에 대한 현실적 우려가 있는 반면, 일부 피험자들은 다감각적 노이즈가 유발하는 인지적 과부하와 업무 환경에서의 피로도 누적에 대한 명확한 한계를 지적하였다.

- 피험자 C: “빠르고 효율적인 업무 처리가 최우선인 고위험 의사결정 환경이나 컨퍼런스룸 등에서는 이러한 다감각적 마찰 자극이 오히려 극심한 인지적 피로도와 업무 방해 요인으로 작용할 우려가 있을 것 같다.”

- 피험자 D: “알고리즘의 불완전성을 경고하려는 설계 취지에는 공감하지만, 자극의 강도가 사용자별로 조율되지 않는다면 장기적인 미디어 소비 환경에서는 피로감이나 귀찮음으로 인해 시스템 제어 옵션을 강제로 꺼버릴 위험이 존재한다.”

이러한 사용자 의견은 제안 프레임워크가 추구하는 매끄러운 인터랙션(Seamful Interaction)이 극단적인 스트레스 자극으로 변질되지 않도록, 사용자의 개인별 감각 임계치에 맞춘 비선형적 스케일링 최적화 연구가 후속 과제에서 필연적으로 전개되어야 함을 보여준다.

4. 성찰적 UX를 위한 하이브리드 미디어 공간 디자인 원칙(Design Principles)

본 연구는 가상 시나리오 분석 및 파일럿 평가를 바탕으로, AI의 편향성과 불확실성 인지를 지원하는 하이브리드 공간 미디어 설계를 위한 세 가지 잠정적 디자인 원칙을 제안한다. 이 원칙들은 탐색적 설계·평가에 기반한 초기 가이드라인으로서, 후속 통제 실험을 통한 검증과 정교화가 필요하다.

1) 원칙 1: 다감각적 마찰의 지각적 우위성(Sensory Friction > Visual Warning)

키네틱 운동 변화와 공간 음향 등 다감각적 변화는 정적 텍스트 경고와 구별되는 지각적 주목을 형성하여, AI 출력의 한계와 신뢰성을 제검토하도록 지원할 가능성이 있다.

2) 원칙 2: 시공간적 의도적 지연을 통한 성찰 조건 형성(Temporal Delay Triggers Reflection)

출력 지연과 시각적 이음새는 AI 답변의 즉각적 수용을 늦추고, 출력 내용과 판단 근거를 추가로 검토할 기회를 제공할 수 있다.

3) 원칙 3: 다감각적 비동기를 통한 기술의 불완전성 가시화(Multimodal Inconsistency Amplifies Distrust)

온라인 시각 정보와 오프라인 공간 자극의 비동기 및 불협화는 편향도·유해성·불확실성 신호를 공간적으로 표현하는 방식으로 활용될 수 있다.

5. 본 연구 아키텍처가 학계에 던지는 시사점

제안된 참조 아키텍처 설계 및 개념적 아티팩트(Conceptual Artifact) 구체화 과정을 통해 본 연구가 획득한 미

디어 공학적 시사점은 다음과 같다. 첫째, III장의 수식 (1), (2), (3)과 같은 정밀한 수학적 파라미터 매핑 모델은 단순한 심미적 연출을 넘어, 추상적 윤리 지표를 제어 가능한 물리 변수로 정량화하는 ‘공학적 번역기(Algorithmic Translator)’로서의 아키텍처적 가치를 지닌다. 둘째, 온·오프라인 플랫폼을 외래 키(Foreign Key) 메커니즘으로 연동한 관계형 데이터 흐름은, 공간의 신체적 자극이 온라인상의 데이터 역추적(Trace)으로 이어지는 ‘성찰적 폐루프(Closed-loop)’ 구조가 하이브리드 미디어 설계의 핵심 요구 사항이 될 수 있음을 시사한다. 이는 본 아키텍처가 특정 도메인에 종속되지 않고 고위험 의사결정 공간이나 리터러시 교육 환경 등 다양한 인터랙션 도메인에 이식될 수 있는 기술적 확장 가능성(Scalability)을 내포하고 있음을 보여준다.

VI. 결론 및 향후 과제

1. 연구 요약 및 결론

본 연구는 매끄러운 상호작용을 중심으로 발전해 온 기존 미디어 인터랙션 패러다임에 대한 대안적 접근으로서, AI 시스템의 편향성과 불확실성을 다감각적 피드백으로 공간에 드러내는 개념적 참조 아키텍처(Conceptual Reference Architecture)인 <프리즘 AI>를 제안하였다. 본 연구는 다감각 자극이 사용자의 인지 과정에 미치는 직접적 인과효과를 실증적으로 입증하기보다, 감각적 마찰이 AI 출력에 대한 관성적 수용을 일시적으로 중단시키고 성찰적 해석을 지원할 수 있는 설계 가능성과 조건을 탐색하였다.

본 연구의 기여는 다음과 같다. 첫째, AI의 신뢰성 및 윤리적 지표를 신체화된 감각 경험으로 전환하는 상위 수준의 인터랙션 방법론으로서 성찰적 UX 프레임워크(Reflective UX Framework)를 제안하였다. 둘째, 실시간 AI 데이터 스트림을 온라인 디지털 UI와 오프라인 실감 미디어 공간 아티팩트에 연동하는 하이브리드 참조 아키텍처의 설계를 제시하였다. 셋째, 매끄러운 디자인과 이음새 있는 인터랙션의 논의를 확장하여, 성찰적 UX를 지원하는 설계 전략으로서 감각적 마찰(Sensory Friction)의 적용 가능성을

논의하였다.

본 참조 모델은 가상 시뮬레이션 기반의 탐색적 사용자 평가(N=35)에 기초하므로, 다감각 마찰의 독립적 효과나 일반화 가능한 인과관계를 검증하는 데 한계가 있다. 그럼에도 본 연구는 향후 성찰적 AI 인터랙션 연구에서 통제 실험, 실제 공간 프로토타입 구현 및 감각 자극 최적화를 위한 개념적·공학적 설계 청사진(Conceptual and Engineering Design Blueprint)으로 활용될 수 있는 초기 설계 근거를 제공한다.

2. 연구의 한계점 및 향후 과제

2.1 가상 시뮬레이션 평가의 한계 및 실물 인프라 구현

본 연구의 사용자 평가는 조건 A·B에 대한 참여자의 회상과 조건 C의 비디오 자극을 비교하는 방식으로 이루어졌다. 따라서 관찰된 평균 차이가 다감각 마찰 자체에 기인한 것인지, 또는 비디오 매체의 생생함과 몰입감의 차이에서 비롯된 것인지를 엄밀히 분리하기 어렵다. 또한 비디오 기반 평가는 실제 공간에서 발생하는 신체화된 감각 경험을 대리하는 방식이므로, 인지적 제동이나 비판적 검토 행동을 직접적으로 검증하는 데 한계를 갖는다.

후속 연구에서는 조건 A·B·C를 동일한 형식의 고충실도 비디오 아티팩트로 제작하여 자극 제시 형식을 통제할 필요가 있다. 나아가 실제 오프라인 공간 프로토타입을 구현한 뒤, 동일 공간에서 키네틱·음향·조명 자극을 소프트웨어적으로 On/Off하는 통제 실험을 수행함으로써 공간적 다감각 마찰의 독립적 효과를 보다 엄밀히 검토해야 한다.

2.2 실험 데이터셋 확장 및 감각 스케일링 알고리즘 고도화

조건별 평균 점수 차이($\Delta=+0.12\sim+0.17$)는 관찰되었으나, 소규모 탐색 평가의 특성상 통계적 유의성 및 일반화 가능한 인과관계를 검증하지 않았다. 따라서 본 연구의 데이터-감각 매핑 수식은 인간의 비판적 인지 전이를 독립적으로 보장하는 모델이 아니라, AI 출력의 편향도·유해성·불확실성을 감각적 신호로 전환하기 위한 초기 설계 가이드라인으로 해석할 필요가 있다. 실제 적용 단계에서는 인지 과부하와 업무 방해 가능성을 고려해야 한다. 후속 연구에서는

사용자별 감각 민감도와 과업 맥락을 반영하여 동적 강도 조절, 개인화된 임계값, 맥락 인식형 활성화 방식을 검토하고, 실제 공간 기반 통제 실험과 표본 확장을 통해 설계 파라미터를 최적화할 필요가 있다.

참 고 문 헌 (References)

- [1] F. Danelid, "Automation Bias in Public Sector Decision Making: a Systematic Review," Master's thesis, Department of Psychology, Umeå University, Umeå, Sweden, 2024. <https://www.diva-portal.org/smash/get/diva2:1870243/FULLTEXT01.pdf>
- [2] R. Choudhuri, C. A. Sanchez, M. Burnett, and A. Sarma, "Thinking Less, Trusting More: GenAI's Impacts on Students' Cognitive Habits," arXiv preprint arXiv:2601.22430, April 2026. doi: <https://doi.org/10.48550/arXiv.2601.22430>
- [3] Y. Ahn, Q. K. Wolter, J. Dick, J. Dick, and Y. R. Lin, "Interactive Counterfactual Exploration of Algorithmic Harms in Recommender Systems," arXiv preprint, arXiv:2409.06916, September 2024. doi: <https://doi.org/10.48550/arXiv.2409.06916>
- [4] K. E. Casey and S. C. da Costa Sousa, "Explainable by Design: A design framework to support the design of explainable user interfaces," in Proceedings of the 36th Annual Conference of the European Association of Cognitive Ergonomics (EACE 2025), New York, NY, USA: ACM, 2025, Art. No. 13. doi: <https://doi.org/10.1145/3746175.3746213>
- [5] J. Kim, "Towards Algorithmic Justice: Human Centered Approaches to Artificial Intelligence Design to Support Fairness and Mitigate Bias in the Financial Services Sector," Senior Thesis, Claremont McKenna College, Claremont, CA, USA, 2024, No. 3498. https://scholarship.claremont.edu/cmc_theses/3498
- [6] M. Chalmers, S. MacColl, and M. Bell, "Seamful design: showing the seams in wearable computing," in Proceedings of the IET Colloquium on Wearable Computing, pp. 11/1 - 11/4, 2000. doi: <https://doi.org/10.1049/ic:20030140>
- [7] U. Ehsan, Q. V. Liao, S. Passi, M. O. Riedl, and H. Daumé III, "Seamful XAI: Operationalizing Seamful Design in Explainable AI," Proceedings of the ACM on Human-Computer Interaction, Vol. 8, No. CSCW1, Art. No. 29, pp. 1 - 32, April 2024. doi: <https://doi.org/10.1145/3637396>
- [8] J. St. B. T. Evans and K. E. Stanovich, "Dual-process theories of higher cognition: Advancing the debate," Perspectives on Psychological Science, Vol. 8, No. 3, pp. 223 - 241, May 2013. doi: <https://doi.org/10.1177/1745691612460685>
- [9] W. Zhang and B. Y. Lim, "Towards Relatable Explainable AI with the Perceptual Process," in Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, Art. No. 382, pp. 1 - 24, April 2022. doi: <https://doi.org/10.1145/3491102.3501826>
- [10] S. Inman, A. Murillo, S. D'Angelo, A. Brown, and C. Green, "Seamful

- AI for Creative Software Engineering: Use in Software Development Workflows,” IEEE Software, Vol. 42, No. 3, pp. 59 – 66, May-June 2025.
doi: <https://doi.org/10.1109/MS.2025.3534085>
- [11] M. Russo and M. E. Vidal, “Leveraging Ontologies to Document Bias in Data,” arXiv preprint, arXiv:2407.00509, June 2024.
doi: <https://doi.org/10.48550/arXiv.2407.00509>
- [12] D. Çay, M. Karyda, M. Minkó, and J. Boros, “How Do Design Students Use Data in Physicalization Courses?” in Proceedings of EKSIG 2025: Data as Experiential Knowledge and Embodied Processes, M. Karyda, D. Çay, Á. K. Bakk, R. Dezső, and J. Hemmings, Eds., Budapest, Hungary: Design Research Society, May 2025, Art. No. 4.
doi: <https://doi.org/10.21606/eksig2025.104>
- [13] J. W. Hwang and S. I. Kim, “Effects of Perceived Structural Layers of Generative AI Interfaces on Creative Experience: Mediating Role of Emotional UX,” Journal of Digital Contents Society, Vol. 27, No. 3, pp. 597 – 611, Mar. 2026.
doi: <https://doi.org/10.9728/dcs.2026.27.3.597>
- [14] S. Wu, “Design Research Case Study of LLM-Based Conversational Sustainability Reflection System for Creative AI Practitioners,” Master’s thesis, KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science (EECS), Stockholm, Sweden, 2025. <http://www.diva-portal.org/smash/record.jsf?pid=diva2:2012476>
- [15] X. Guan, N. Demchak, S. Gupta, Z. Wang, E. Ertekin Jr., A. Koshiyama, E. Kazim, and Z. Wu, “SAGED: A holistic bias-benchmarking pipeline for language models with customisable fairness calibration,” in Proceedings of the 31st International Conference on Computational Linguistics, Abu Dhabi, UAE: Association for Computational Linguistics, pp. 3002 – 3026, January 2025. <https://aclanthology.org/2025.coling-main.202/>
- [16] M. D. Muralikumar, Y. S. Yang, and D. W. McDonald, “A Human-Centered Evaluation of a Toxicity Detection API: Testing Transferability and Unpacking Latent Attributes,” ACM Transactions on Social Computing, Vol. 6, No. 1 – 2, Art. No. 4, pp. 1 – 38, 2023.
doi: <https://doi.org/10.1145/3582568>
- [17] H. Chen, Y. Ji, and D. Evans, “Do Prevalent Bias Metrics Capture Allocational Harms from LLMs?,” in Proceedings of the 3rd Workshop on Insights from Negative Results in NLP, Association for Computational Linguistics, pp. 44–57, January 2025.
doi: <https://doi.org/10.18653/v1/2025.insights-1.5>
- [18] Y. Geifman and R. El-Yaniv, “Selective Classification for Deep Neural Networks,” in Advances in Neural Information Processing Systems (NIPS 2017), Vol. 30, pp. 4885–4894, December 2017.
doi: <https://doi.org/10.48550/arXiv.1705.08500>
- [19] R. Plomp and W. J. M. Levelt, “Tonal consonance and critical bandwidth,” The Journal of the Acoustical Society of America, Vol. 38, No. 4, pp. 548–560, 1965.
doi: <https://doi.org/10.1121/1.1909741>
- [20] C. S. Herrmann, “Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena,” Experimental Brain Research, Vol. 137, No. 3–4, pp. 346–353, 2001.
doi: <https://doi.org/10.1007/s002210100682>
- [21] T. Cao, F. Wan, C. M. Wong, J. N. Da Cruz, and Y. Hu, “Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces,” BioMedical Engineering OnLine, Vol. 13, Art. No. 28, pp. 1–13, 2014.
doi: <https://doi.org/10.1186/1475-925X-13-28>
- [22] J. Zimmerman, J. Forlizzi, and S. Evenson, “Research through design as a method for interaction design research in HCI,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA: ACM, pp. 493 – 502, April 2007.
doi: <https://doi.org/10.1145/1240624.1240704>
- [23] M. Vered, T. Livni, P. D. L. Howe, T. Miller, and L. Sonenberg, “The effects of explanations on automation bias,” Artificial Intelligence, Vol. 322, Art. No. 103952, pp. 1 – 24, Sept. 2023.
doi: <https://doi.org/10.1016/j.artint.2023.103952>

저 자 소 개



강 자 원

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

저 자 소 개



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), 디지털콘텐츠(Digital Contents) 등