

[Technical White Paper]

ACAIP: 엣지 AI 기반 아동 행동 인지 및 실시간 인터랙티브 플랫폼

버전: 1.0 (2026. 02) 분야: Edge AI, Action Anticipation, Child Safety

1. 개요 (Executive Summary)

본 프로젝트는 아동의 돌발 행동으로 인한 사고를 예방하기 위해 **0.5 초 선제적 예측(Proactive Anticipation)** 기술을 구현한다. NVIDIA의 차세대 Blackwell GB10 인프라를 활용하여, 기존 클라우드 기반 AI의 한계인 지연 시간(Latency) 문제를 극복하고 엣지(Edge) 환경에서 실시간 추론 및 시각적 피드백을 제공하는 시스템 구축을 목표로 한다.

- 플랫폼 명칭 : Agentic Child-Action Interaction Platform (ACAIP)
- 특정 시나리오에 종속되지 않는 '**범용 행동-콘텐츠 동기화 엔진**'. 아동의 멀티모달(Vision, Voice, Pose) 데이터를 실시간 수집하여 표준화된 '상호작용 프로토콜'로 변환, 외부 IP 콘텐츠를 실시간 제어하는 기술

2. 개발배경 및 필요성

2.1. 콘텐츠 경험의 한계: "단조로움과 몰입감 저해"

기존의 아동용 인터랙티브 시스템은 아이의 움직임과 상관없이 정해진 영상을 반복 재생하는 경우가 많습니다.

- 문제점 (Fixed Pattern):** 기존 시스템은 아이가 어디에 있는 똑같은 애니메이션을 보여주는 '수동적 재생' 구조입니다. 아이들은 금방 패턴을 파악하고 흥미를 잃습니다.
- 해결 근거 (Behavior-Content Mapping):** 아동의 ****동작 벡터(방향/속도)****와 ****감정 지표(표정/소리)****를 분석하여 실시간으로 시나리오를 분기시킴으로써, 아이가 자신의 행동으로 세상을 변화시킨다는 강력한 '효능감'과 '몰입감'을 제공해야 합니다.

2.2. 인터랙션 지연의 문제: "인지 부조화 발생"

인간, 특히 감각이 예민한 아동은 자신의 행동과 시스템의 반응 사이에 미세한 시차가 발생하면 본능적으로 이질감을 느낍니다.

- **문제점 (Latency Mismatch):** 센서가 동작을 인식하고 서버를 거쳐 프로젝터가 반응하는 데 걸리는 시간(Latency) 때문에, 아이의 동작과 콘텐츠 반응 사이에 '엇박자'가 생깁니다. 이는 상호작용의 질을 급격히 떨어뜨립니다.
- **해결 근거 (0.5s Predictive Agent):** 아이가 움직인 **후**에 반응하는 것이 아니라, 동작 벡터를 분석해 0.5 초 뒤의 위치를 ****미리 예측(Predictive)****하여 콘텐츠를 마중 나가게 함으로써 지연 시간을 기술적으로 '제로(0)'화하는 경험이 필요합니다.

2.3. 안전 관리의 한계: "사후 약방문식 대응"

공공 놀이 시설이나 체험형 전시관에서 안전 사고는 순식간에 일어납니다.

- **문제점 (Reactive Safety):** 현재 대부분의 안전 시스템은 사고가 발생한 지점에서 경보를 울리는 '반응형'입니다. 하지만 아이들의 돌발 행동은 사고 직전의 찰나에 제어되어야 합니다.
- **해결 근거 (Trigger Logic):** 아동의 동작 벡터가 위험 구역이나 위험 속도로 진입할 것을 0.5 초 전에 미리 예측하여, 해당 지점에 ****즉각적인 시나리오 트리거(시각적 차단막, 경고 그래픽 등)****를 발생시켜 사고를 물리적으로 예방해야 합니다.

3. TOTI 기술 솔루션 (Technical Solution)

3.1 하드웨어 아키텍처: NVIDIA Blackwell GB10

- **Dual-Node Clustering:** 2 대의 GB10 을 NVLink 로 결합하여 256GB 의 통합 메모리 풀(Unified Memory Pool)을 확보한다.
- **FP4 Acceleration:** Blackwell 의 5 세대 Tensor Core 를 활용하여 모델 성능 저하 없이 연산 속도를 극대화한다.
- **Multi-Sensor Fusion:** 6 개의 LiDAR 와 6 개의 Depth 카메라에서 유입되는 방대한 점구름(Point Cloud) 데이터를 실시간으로 병렬 전처리한다.

3.2 알고리즘: ST-GNN (Spatio-Temporal Graph Neural Network)

- **Graph Representation:** 아이의 주요 관절 17 개를 노드(Node)로, 연결을 엣지(Edge)로 구성하여 신체 구조적 특징을 학습한다.
- **Temporal Attention:** 이전 프레임의 움직임 패턴을 분석하여 향후 0.5 초(30 프레임 내외)의 궤적을 확률적으로 계산한다.
- **Action Anticipation:** 단순한 위치 추적을 넘어 '뛰기', '넘어짐', '손 뺀기' 등의 의도를 사전에 분류한다.

4. 시스템 워크플로우 (Workflow)

1. **Sensing:** 6 개 지점에서 아동의 3D 공간 데이터 수집.
2. **Preprocessing:** GB10 Node 1 에서 노이즈 제거 및 스켈레톤 추출.
3. **Inference:** GB10 Node 2 에서 ST-GNN 모델을 통한 0.5 초 뒤 좌표 예측.
4. **Interaction:** 예측된 좌표에 6 대의 프로젝터가 실시간으로 안전 가이드(바닥 조명, 경고 표식) 투사.
5. **Feedback Loop:** 실제 움직임과 예측값의 오차를 실시간 학습(On-device Learning)에 반영.

5. 정량적 목표 (KPI)

- **End-to-End Latency:** 200ms 이하 (센싱부터 투사까지)
- **Prediction Horizon:** 500ms (0.5 초 앞선 예측)
- **Accuracy:** 아동 돌발 행동 예측 정확도 92% 이상
- **Concurrent Tracking:** 최대 10 명의 아동 동시 추적 및 개별 가이드 제공