

대표성 편향 (Representation Bias)

학습 데이터가 실제 인구 분포를 제대로 반영하지 못할 때 발생하는 편향



데이터 불균형 문제

- 특정 인종/민족 과소 대표
- 성별 불균형
- 연령 분포 왜곡
- 희귀 질환 데이터 부족
- 지리적 편중



결과적 영향

- 소수 집단에서 낮은 정확도
- 오진율 증가
- 치료 권장 편향
- 건강 격차 심화
- 신뢰도 저하



보정 전략

재샘플링

Over/Under sampling, SMOTE

가중치 조정

Class weights, Sample weights

데이터 증강

Augmentation, Synthetic data