

KNIH-KSBI Joint Healthcare AI Challenge 2026

멀티모달 코호트 합성
데이터를 활용한
인공지능 모델 개발

주최: 질병관리청 국립보건연구원(KNIH), 한국생명정보학회(KSBI)
연계: BIOINFO/GIW ISCB-Asia 2026



2026. 9. 8. 사전 설명회

헬스케어 AI의 새로운 패러다임: 멀티모달 데이터의 융합



과거-현재: 단일 모달리티의 한계

단일 영상이나 유전체만으로는 복잡한 만성질환 및 노화의 궤적을 온전히 설명할 수 없음.

미래: 멀티모달 융합 (Multimodal Integration)

임상역학, 의료영상, 유전체 데이터가 얹힌 복합적 인과관계를 입체적으로 학습하는 AI 모델의 부상.

비전: KNIH의 목표

한국인 특화 다차원 데이터를 통한 보건의료 AI 리터러시 향상 및 헬스케어 특화 데이터 사이언티스트 양성.

KNIH-KSBI 공동주관: 당신의 모델을 증명할 무대



Who (참가 대상)

- 대학생, 대학원 석·박사 과정 재학생 또는 휴학생 (졸업 예정자 포함)
- 관련 분야 연구원 등 개인 또는 팀 (최대 3명)



What (핵심 미션)

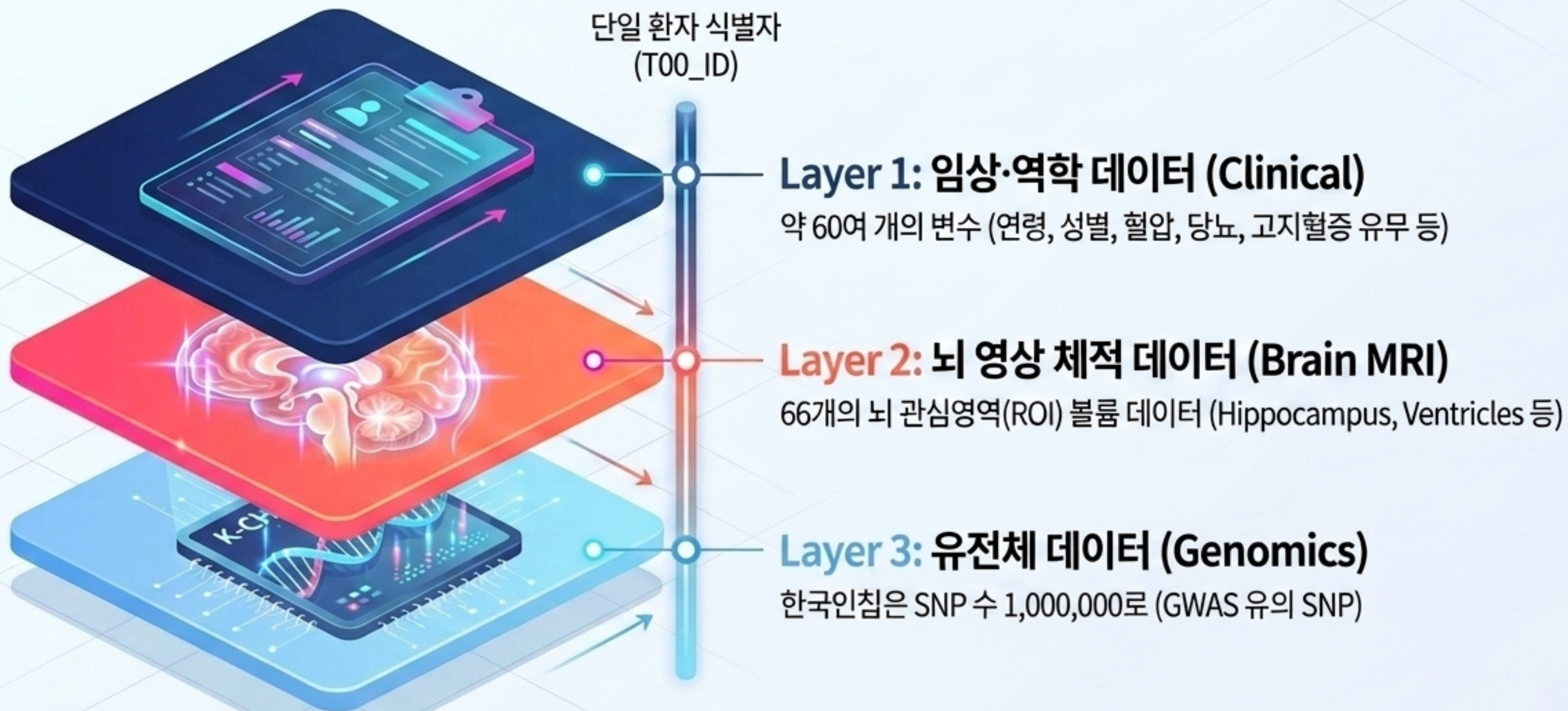
- 국립보건연구원 제공 'KoGES 코호트 기반 멀티모달 합성 데이터' 활용
- 질병 예측, 뇌 나이 추정 등 헬스케어 특화 인공지능 모델 개발 및 검증



Where (결선의 무대)

- 아시아 최대 생명정보학회 BIOINFO/GIW
ISCB-Asia 2026 연계
- 11월 17일, BIOINFO/GIW 학술대회장 발표,
11월 20일 시상

멀티모달 퍼즐의 완성: 제공 데이터 오버뷰



⚠ 모든 데이터는 철저히 **가명 및 합성 처리**된(Synthetic) 교육용 데이터입니다.

데이터 아키텍처: 단면(Cross-sectional)과 추적(Longitudinal)

Stage 1: BASE DATA (횡단 분석용)



- 규모: 10,000명 (기반조사 단면)
- 특징: 풍부한 표본 수를 활용한 질병 분류, 연령대별 뇌 볼륨 분포, 멀티모달 상호작용 횡단 분석에 최적화.

⌚ 약 10.8년 경과

Stage 2: FOLLOW_01 ~ FOLLOW_05 (종단 분석용)



- 규모: 1,000명 (기반) → 698명 (4차 추적 완료)
- 특징: 유전형(DNA)은 불변하나, 임상 변수 악화 및 뇌 위축(연령 가증치) 발생. 질환의 신규 발생 예측 및 노화 궤적 시계열 모델링에 최적.

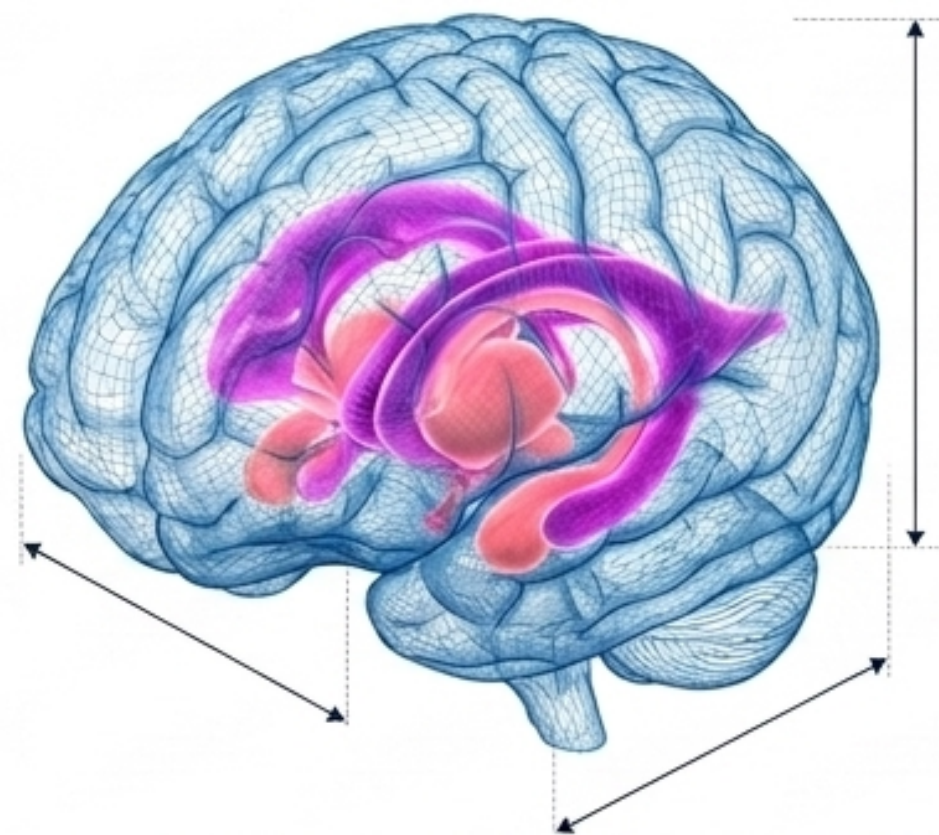
Modality 1 & 2: 임상 패턴과 뇌 구조적 변화의 맵핑

Clinical Data (임상·역학)



- 주요 변수: 나이(T_AGE), 성별(T_SEX), 수축기혈압(T_SBP), 고혈압(T_HTN), 당뇨(T_DM), 고지혈증(T_LIP)
- 모델링 포인트: 결측치(99999 등)의 적절한 전처리 및 질환 유무에 따른 이진 변환(Binary encoding)이 성능의 핵심 관건.

Brain MRI Volumetric Data (뇌영상 체적)



- 66개 ROI 체적: 문헌 기반 생물학적 패턴 내재 합성 데이터
- 주요 측정치: 총 두개강 용적(ICV), 전체 뇌(Whole Brain), 해마(Hippocampus), 편도체(Amygdala), 뇌실(Ventricles)
- 특징: 45세 남성 기준 기댓값에서 연령, 성별, 질환 유무에 따른 개인별 변동성이 정교하게 합성됨.

Modality 3: 한국인 특화 유전체 정보 (K-CHIP)

한국인 특이 유전 변이 정보

- 서양인 데이터와 차별화되는 한국인 코호트만의 유전적 특성값 포함
- 임상정보의 질환 정보(당뇨, 고혈압, 고지혈증) 등과 연관된 유전자 포함

Playbook: 멀티모달 모델링 전략 및 활용 방향

추천: RF / XGBoost

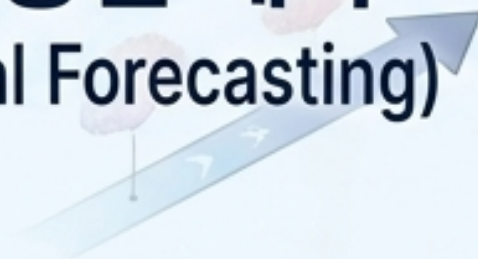
뇌 나이 추정 모델 (Brain Age Gap 예측)



66개 뇌 ROI 볼륨과 임상 변수를 융합하여 환자의 '생물학적 뇌 나이'를 예측하는 회귀 기반 앙상블 모델 개발.

추천: LSTM / LMM

만성질환 종단 예측 (Longitudinal Forecasting)



FOLLOW 데이터를 활용해 초기(Base) 데이터만으로 10년 후의 고혈압 신규 발생이나 해마 위축량을 시계열 예측.

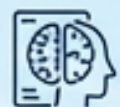
추천: SVM / LightGBM

다중오믹스 기반 위험 스코어 (치매 고위험군 조기 분류)



APOE 유전형, 질환별 PRS, 그리고 임상 데이터를 결합하여 고위험군을 조기에 판별해내는 복합 위험 분류기 개발.

How to Win: 심사 및 평가 기준



활용 가능성 (Applicability)

개발된 모델이 실제 헬스케어 현장이나
후속 임상 연구에 적용될 수 있는
임상적/생물학적 해석 가능성을 갖추었는가?



데이터 활용 (Data Integration)

임상, 영상, 유전체라는 성격이 다른
멀티모달 데이터를 얼마나 논리적이고
효과적으로 융합(Fusion)하여 분석하였는가?



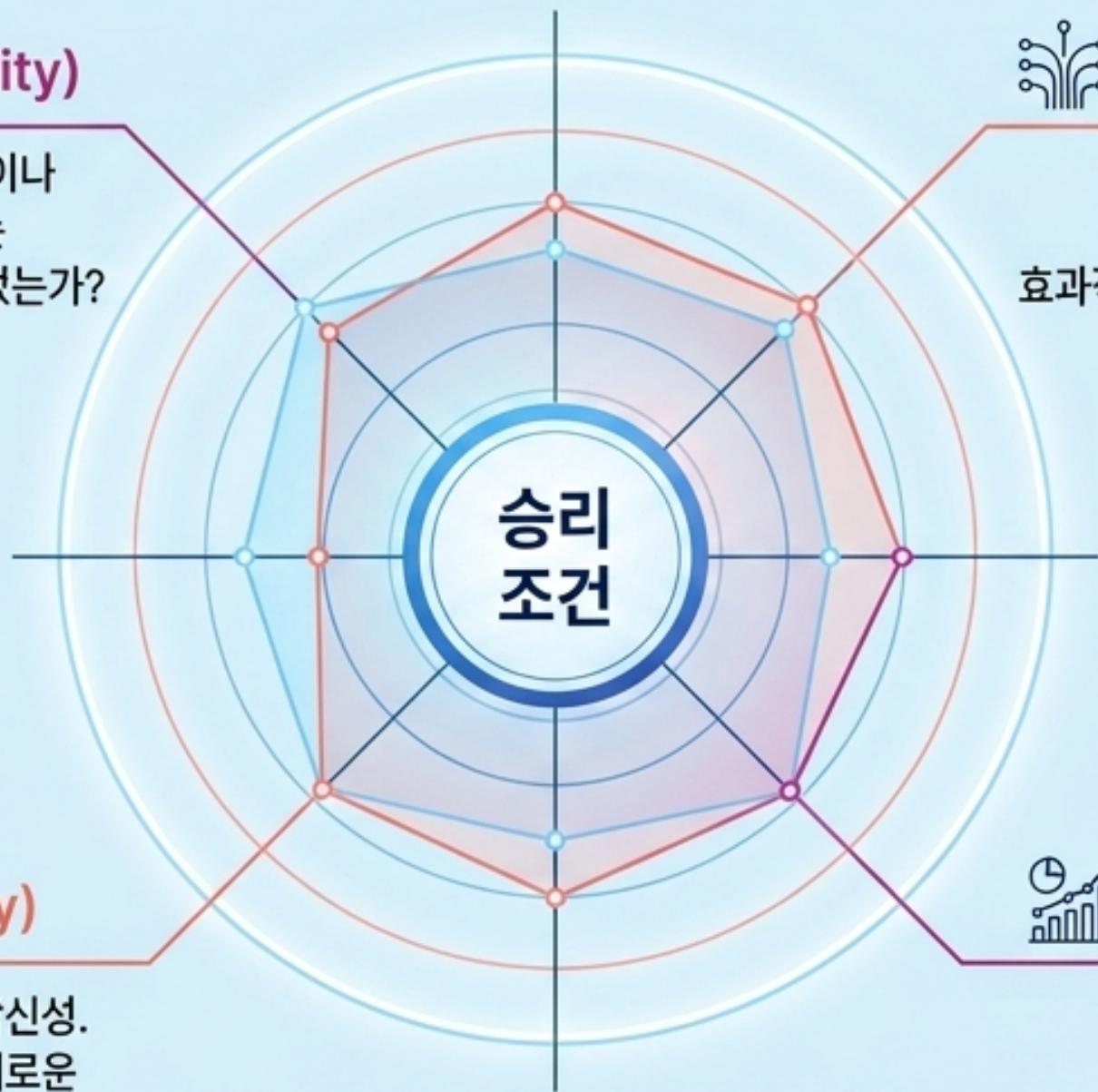
독창성 (Originality)

문제 접근 방식과 가설 설정의 참신성.
기존의 단순 회귀를 넘어서는 새로운
AI/ML 아키텍처 제안.



성능 (Model Performance)

예측 및 분류 모델의 통계적 타당성,
교차 검증(Cross-validation), 객관적인
평가지표(R^2 , AUC 등)의 우수성.



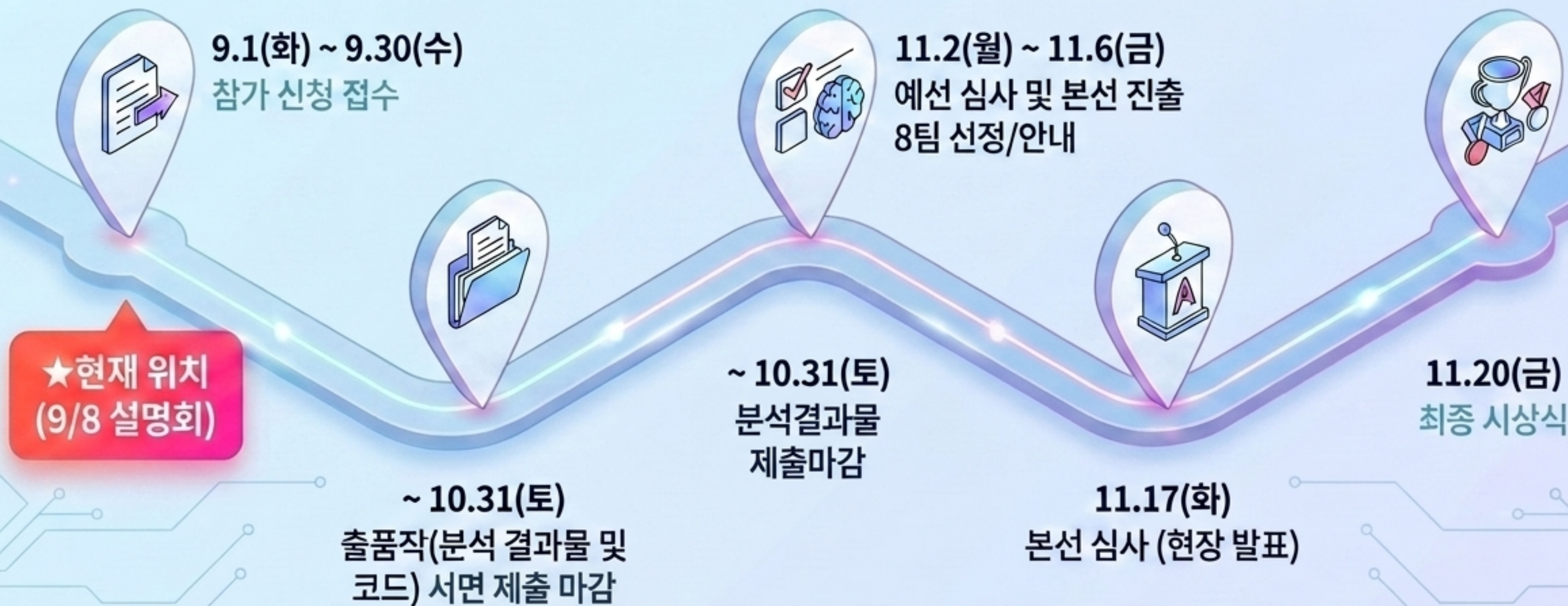
The Reward: 총상금 500만 원 및 8개 팀 시상



※ 수상팀 특전: 아시아 최대 생명정보학회(BIOINFO/GIW 2026) 본선 무대에서 연구 결과물 특별 발표 기회 제공

Milestones: 본선을 향한 로드맵

일정은 주관 기관 사정에 따라 변경 가능합니다.



Mission Briefing Complete.

도전을 시작하십시오.

지금 바로 한국생명정보학회 누리집에서 세부 규정을 확인하고 팀을 구성하세요.
참가 신청, 결과 제출 및 모든 문의는 공식 이메일을 통해서만 진행됩니다.



[대회 공식 누리집]



[참가 신청 안내]

공식 이메일: healthcareai.knih@gmail.com
국립보건연구원 (nih.go.kr) / 한국생명정보학회 (ksbi.or.kr)

멀티모달 헬스케어 데이터에 숨겨진 진실을 밝혀낼 여러분의 혁신적인 AI 모델을 기대합니다.