

## 별첨1-4 [제출물 양식] AI 활용 보고서 (Track1)

※ (제출 시, 삭제) 제출된 자료(논문 및 AI Agent를 제외한 자료, 예: AI 활용보고서, 활용데이터 목록 등)는 심사 목적으로만 사용될 예정이며, 외부로 반출되지 않습니다.

### 1. 참가팀 정보

팀 정보	팀 이름	EGO-Evolution	참가자	허의학
연구 내용	지정/선택 여부	지정/선택	연구분야	AutoML / Tabular ML / Multi-objective Optimization / Reproducibility
	연구주제	Constraint-aware Evolutionary AutoML with Verifiable Provenance (AUC 최대화 + 제약 만족)		

### 2. 활용 AI 모델 정보[예시]

AI 모델명	GPT-5.2	AI 모델 URL	https://chatgpt.com/

※ 자체제작 AI 모델 활용 시, 재현성 검증을 위해 해당 AI 모델을 사용 또는 다운받을 수 있는 URL을 제시

### 3. 연구 절차별 AI 기여도(자체평가)

연 번	연구 절차	배점 (A)	AI 기여도 자체평가 (0~100%) (B)	AI 기여도 산출 (C=A × B)	AI 작업내용 (간략히)
1	주제 선정 및 연구문제 도출	10	60	6	Track1 요구사항/평가지표(AUC, CSR) 기준으로 "진화탐색+제약+검증+LL M 힌트" 프레임을 설계
2	선행연구 조사 및 문헌 검토	5	60	3	AutoML/NSGA-II/제약 최적화/모델 검증 루틴 관련 키워드 정리 및 방법 비교
3	연구목적 및 문제 정의	15	55	8.25	"성능(AUC) vs 제약(CSR)"의 다목적 최적화 문제로 명시, 실험 설계/지표 정의 문장화
4	연구계획 및 방법론 설정	20	70	14.0	genome 설계, mutate/crossover, Pareto front, 검증 단계, baseline/ablation suite 구성안 작성
5	자료(데이터) 수집	10	50	5.0	OpenML/Sklearn 데이터 로딩, 비활성 dataset 대응(버전/타겟컬럼 추론) 예외처리 가이드
6	자료(데이터) 분석	20	60	12.0	suite 결과 요약 스크립트, mean±std 산출, 결과표/해석 포인트 정리
7	논문 작성	5	50	2.5	템플릿(NeurlPS 스타일) 구조에 맞춰 섹션 구성/표/문장(영문) 초안 작성
8	자체 리뷰 및 수정	15	60	9	에러 로그 기반 원인 분석(경로/요약파일 누락), 실행 재현성 체크리스트 작성
총점		100		60.25	

※ AI 기여도 자체평가 기준: "의사결정 및 행동의 주체가 누구인가"

※ 항목별 배점 x 기여도 자체평가의 총합이 60% 이상이어야 함

**붙임**

**AI 작업내용 상세 (활용 증빙 스크린샷, URL, 로그파일 등)**

**1. 주제 선정 및 연구문제 도출**

Track1 목표를 "제약(학습시간/특징수 등)을 만족하면서 AUC를 최대화"로 재정의하고, 이를 다목적 최적화(AUC ↑, CSR ↑)로 구조화함.  
 모델 후보(로지스틱/랜포/GBDT/SVM) + 특징 선택(k-best) 조합을 genome으로 두고 진화탐색으로 자동 선택하는 방향을 결정함.

## 2. 선행연구 조사 및 문헌 검토

AutoML/하이퍼파라미터 탐색(Random/EA), Pareto front, 제약 최적화(CSR) 개념을 정리해 본 과제 구조에 맞게 재해석함.  
"검증(verification)"을 단순 성능 보고가 아닌 재현성·무결성(로그/설정/결과 파일) 확보 단계로 포함시키는 구성을 도출함.

### 3. 연구목적 및 문제 정의

목적: "도메인별 탭уляр 이진 분류에서, 주어진 제약을 만족하면서 AUC를 극대화하는 자동 파이프라인"  
 문제: 후보 모델과 특징수/하이퍼파라미터 공간이 넓어 수동 탐색이 비효율적 → 진화탐색으로 효율화.

**4. 연구계획 및 방법론 설정**

(1) 데이터 로딩 → (2) train/val split → (3) EA로 candidate 평가(AUC/CSR) → (4) Pareto front 추출 → (5) best genome 재학습 후 test 평가.  
ablation(LLM 제거/제약 제거/검증 제거/단일목표) + baseline(각 단일 모델 고정) + random search를 suite로 구성함.

## 5. 자료(데이터) 수집

OpenML 데이터는 “비활성(inactive) 버전/타겟 메타데이터 누락” 가능성이 있어, 이름 기반 fetch 및 target column 추론 로직을 포함하도록 설계함.  
바이오 도메인은 오프라인 재현성을 위해 sklearn breast cancer 데이터도 함께 사용 가능하도록 함.

**6. 자료(데이터) 분석**

각 seed/실험 폴더의 metrics\_best.json을 수집해 summary.csv, summary\_mean\_std.csv, summary\_table.md로 집계.  
핵심 비교 지표는 test AUC(mean±std), test acc(mean±std)로 통일하여 표로 제시함.



**7. 논문 작성**

템플릿(2026 AI Co-Scientist Challenge Korea/NeurIPS 스타일)에 맞춰 Abstract-Method-Experiments-Results-Limitations-Broader Impact-Checklist 순서로 작성함.  
결과표는 suite 집계표를 그대로 LaTeX table로 옮겨 제출용으로 작성함.

**8. 자체 리뷰 및 수정**

실행 중 경로/요약파일 누락 오류(FileNotFoundError) 원인을 "suite\_root 자동탐색/출력 폴더 규칙 불일치"로 정리하고, suite\_summary가 항상 suite 폴더 내부에 summary\*.csv를 생성하도록 수정함.  
스크립트/설정/로그/결과파일이 동일 구조로 저장되도록 통합 실행(run\_suite→suite\_summary→plot) 동선을 점검함.