



## 대회 매뉴얼 (제공용)

주최  
**SNUH** 서울대학교병원

주관  
**OUER** Lab

**SNUH** Venture

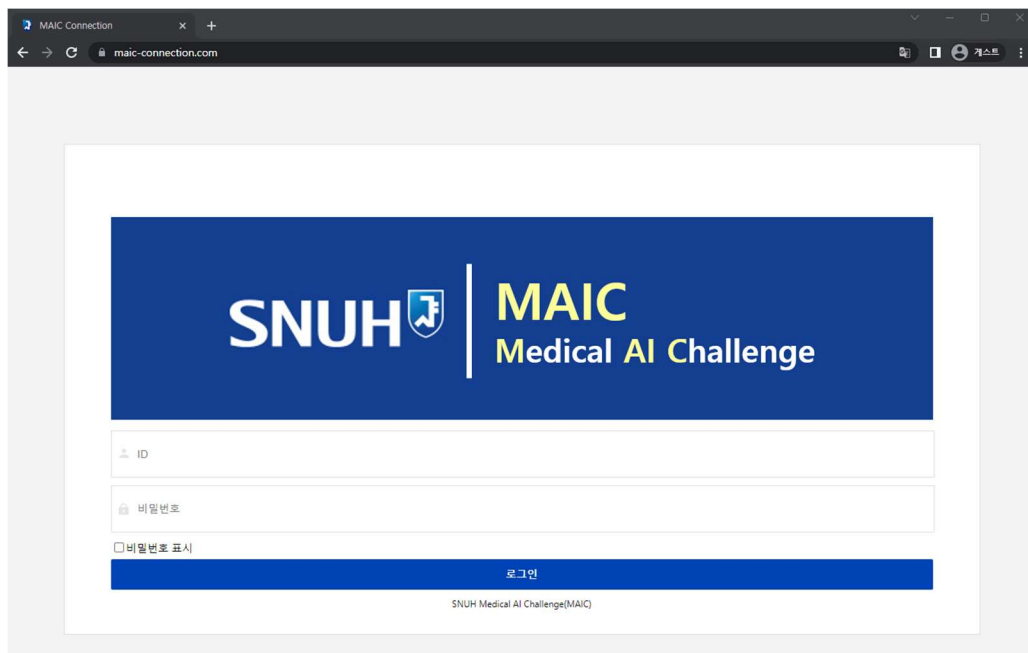
후원  
**NIA** 한국지능정보사회진흥원

1. 주요 사용방법.....	3
A. 대회 환경 접속방법.....	3
B. 중요 부정행위 관련 공지 .....	4
C. 대회 환경 사용시 주요 주의사항 .....	5
2. 대회 개요 및 사업설명 .....	7
A. 대회 개요 .....	7
B. 추진 배경 .....	7
C. 추진 목표 .....	8
D. 대회 규정 .....	8
E. 평가규칙.....	11
F. 게시판 사용규칙 안내 .....	13
3. 환경 구성 설명 .....	15
A. 기본 제공 사양.....	15
B. 기본 Anaconda 가상환경 제공.....	15

# 1. 주요 사용방법

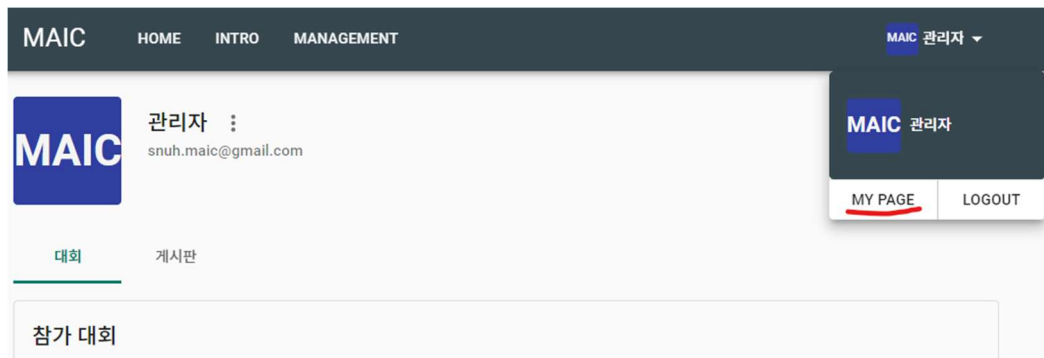
## A. 대회 환경 접속방법

- 노트북 접속 페이지에서 아이디와 비밀번호를 입력하면 참가한 팀의 주피터 노트북으로 이동합니다.



- 노트북 접속 페이지는 주소창에 maic-connection.com 입력 또는 maic.or.kr 홈페이지에서 아래의 방법으로 접속 가능합니다.
- 홈페이지 상의 참가 대회의 배너를 클릭 또는 마이페이지의 참가한 대회로 이동하면, 노트북 접속 버튼이 있습니다.





- 노트북 접속 버튼 클릭을 통해 주피터 노트북 접속 페이지로 이동합니다.



## B. 중요 부정행위 관련 공지

- 데이터 반출 시도는 절대 금지합니다.(적발시 바로 탈락)
- 수상자들은 기한까지 기술 문서를 작성하여 제출합니다.
- 저장공간을 5Tb 이상 사용하지 말아주세요.
  - 5TB 이상 사용시 경고
  - 7TB 이상 사용시 실격

## C. 대회 환경 사용시 주요 주의사항

✧ **dataset** 폴더는 **읽기전용**으로 제공됩니다.

- 해당 폴더에서 업로드와 저장이 제한됩니다.
- dataset 내부의 데이터 변경 시 동시 제공완료 후 공지 드립니다.
- dataset 폴더 내부의 데이터를 폴더 외부로 복사하시면 안됩니다

✧ 할당된 **메모리**는 **100GB** 로 제한합니다.

- Memory 초과로 인한 상황(노트북 OOM 커널 종료 등)에 대해서는 확인 해 드리지 않습니다.
- 사용 하지 않는 노트북 파일들이 Running 상태일경우 사용하지 않는 메모리의 점유가 늘어납니다. 노트북의 tree 에서 사용하지 않는 노트북 파일들의 Running 상태를 체크해주세요.

✧ Cuda, Nvidia Driver, OS(ubuntu)는 **버전**이 **고정**됩니다.

- Cuda : 11.7
- Nvidia Driver : 515.65.01
- OS : Ubuntu 20.04

✧ **Empty 환경** 구성 전에 기본 OS, Cuda, Nvidia Driver 에서 **버전에 문제**가 없는지 꼭 한번 **확인** 해 주세요

✧ **Vanila 환경**의 경우 제공된 **버전**을 **수정**하실 경우  
**에러사항**에 대해 **지원해** **드리지** **않습니다**.

- Python, Torch, Tensorflow 등
- 커스텀 환경 및 버전이 필요하신 분들은 empty-maic2022 환경을 사용하시기 바랍니다.

✧ 시작부터 종료까지 특정 기간에 몰아서 태스크를 진행하시기  
보다는, **꾸준히 참여**해주시는 것이 대회 참여 모니터링에  
있어 **불이익**을 받지 **않을** 수 **있으니** **적극 참여** 부탁드립니다.

- 대회 상황에 따라 진행 중 제출 횟수 제한 정책이 업데이트 될 수 있으니 참고 및 양해 부탁드립니다.

✧ **작업이 없을 경우** 참가 **자격을 잃을** 수 있습니다.

- 대회 시작 후 일정이상 자원 사용 및 작업이 없을경우 대회 참여 의사가 없는 것으로 간주 될 수 있습니다.
- Hardware Stress Test 프로그램들의 사용을 엄격히 금지합니다.

✧ 대회에 필요한 **자료**는 **dataset** 폴더의 다음 폴더 및 파일  
들을 **참조** 부탁드립니다.

- 1\_Train\_Val : 학습용 데이터 300 개
- 2\_Test : 테스트용(채점용)데이터 102 개
- version-check-code.ipynb : 대회 환경 버전 확인

## 2. 대회 개요 및 사업설명

### A. 대회 개요

- 수면 데이터는 인생의 1/3 을 차지하고 있는 주요 라이프 로그 데이터로서, 의료 분야는 물론 산업계에서도 관심이 많은 주요 빅데이터 입니다. 지난 2021 년 2 월과 10 월에는 각각 수면 질 데이터(수면다원검사 이미지 데이터)와 적외선 수면 동영상 데이터를 활용하여 수면 단계를 추론하는 국내 최초의 수면 인공지능 경진대회(Sleep AI Challenge ver. 1 및 2)가 성공적으로 개최되었으며, 많은 팀들이 참여하여 우수한 기량을 보여주었습니다. 3 회차를 맞는 이번 경진대회에서는 일상 생활에서 수집 가능한 액티그래피, 심전도 및 소리 데이터 등의 자가 수면 검사 데이터를 이용하여 가장 흔한 수면 질환인 수면무호흡증 리스크를 예측하는 모델링이 태스크로 주어집니다. 자연지능(인간)의 판독 과정을 모사한 새로운 인공지능 학습 전략으로 교내의 인공지능 분야 연구자들이 서로 배울 수 있는 유익한 공유의 장이 될 수 있도록 노력하겠습니다.

### B. 추진 배경

- 사업명 : 2022 년 인공지능 학습용 데이터 구축 사업(한국지능정보사회진흥원)
- 과제명 : 자가 검사 데이터 구축 사업  
(총괄책임자: 서울대학교병원 신현우 교수)

### C. 추진 목표

- 수면 데이터는 인생의 1/3 을 차지하는 주요 라이프 로그 데이터!
- 자가 수면 검사 데이터(움직임, 심전도, 호흡음)를 이용하여 수면무호흡증의 리스크 예측 모델링에 도전하라!

### D. 대회 규정

- 개발 환경 및 프레임워크
  - 부문별 참가팀별로 GPU 도커 컨테이너 (Ubuntu 20.04, Anaconda) 인스턴스가 제공되며, Jupyter notebook 플랫폼에서 패키지 설치 및 코딩을 할 수 있습니다.
  - Python 언어만을 사용할 수 있습니다.
  - 데이터 반출은 불가능하며, 이를 방지하기 위하여 플랫폼상에 출력이 제한됩니다. 데이터를 무단으로 반출하거나 이를 시도하는 행위는 부정행위에 해당합니다.
  - 사후 검증을 위하여 가독성이 좋은 코드 작성이 적극 권장됩니다. (성능에 대한 재현성 검증 불가시 수상에서 제외될 수 있습니다.)



- 데이터 및 모델 규정

- 주최측이 대회 플랫폼을 통해 제공한 데이터 외의 별도의 데이터를 사용할 수 없습니다.
- TensorFlow 와 PyTorch 패키지에서 기본적으로 제공하는 것에 한해 사전 훈련된 모델 (Pre-trained model)을 사용 가능합니다.

- 제출 및 평가

- 참가자는 대회 종료 시각 전까지 추론 결과가 담긴 파일을 채점 서버에 제출할 수 있습니다.
- 파일 양식을 반드시 준수하여 전송하여야 하며, 주최측은 잘못 구성된 파일로 인한 채점 오류에 대하여 책임을 지지 않습니다.
- 평가 데이터에 편향되게 코드 및 인공지능 모델을 구성하는 행위는 부정 행위에 해당합니다.
- 본 대회 시작(예비 기간을 제외함)이 경과한지 2 일 이상 GPU 사용량이 1% 미만이거나 3 일 이상 결과 미제출 등 현저하게 참여 의사가 없다고 판단되는 경우, 해당 팀을 대회에서 제외하고 차순위 팀을 참여시킬 수 있습니다.

☆ 대회 참가에서 결과 제출까지 많은 시간이 걸릴 것이 예상되어도 주기적으로 결과 제출을 하실 것을 권장드립니다.

- 결과물 제출 및 재현성 검증

- 공정성을 위해 재현성 검증을 통하여

- ◆ 인공지능 알고리즘이 일관된 추론을 하는지

- ◆ 평가 데이터에 편향되게 학습하는 등 부정행위가 이루어지지 않았는지를 사후 검증하고 있습니다.

- 사후 검증 참여 및 협조는 모든 참가자들의 의무이며 이를 이행하지 못할 경우 수상팀에서 제외됩니다.

- 대회가 종료되기 전 형식에 맞추어 가독성 있게 메인 코드 (main.py)를 작성하여 대회 플랫폼의 제출 폴더에 위치시켜야 합니다.

- 사후 검증 과정에서 메인 코드에 오류가 발생할 경우, 사소한 문제에 한하여 주최측은 참가자들에 소명을 요청할 수 있습니다. 참가팀이 이에 2 일 이내 응답하지 않거나 중대한 문제일 경우 수상팀에서 제외됩니다.

- 사후 검증 과정에서 제출 결과와 참가자의 코드를 직접 실행하여 산출한 평가 지표가 1% 이상 차이나는 경우 순위가 변경될 수 있습니다. 매 시행에서 1% 이상 차이가 나는 등 일관된 결과가 반환되지 않을 경우 수상팀에서 제외됩니다.

- 기술 문서 작성

- 수상자들은 작성 요령을 준수하여 추후 공지되는 기한까지 기술 문서를 작성하여 제출하여야 합니다. 기술문서를 제출하지 않은 경우 수상이 취소되고 차순위 팀으로 순위가 넘어갑니다.

## E. 평가규칙

- 정량평가

- 수면무호흡증 리스크 유무 분류 (Binary Classification)
- Measurement :  $F_1$  Score
- 본 경진대회에서 주어지는 Task 는 수면무호흡증 리스크의 유무를 분류하는 이진 분류 문제이지만, 각 클래스별  $F_1$  Score 를 이용해 계산된 Macro  $F_1$  Score 를 이용해 모델 성능을 평가합니다.
- 구체적으로는, 수면무호흡증 리스크가 없는 Normal 클래스에 대한  $F_1$  Score( $F_{1,Normal}$ )와 OSA 리스크가 있는 OSA 클래스에 대한  $F_1$  Score( $F_{1,OSA}$ )의 산술 평균인

$$\frac{F_{1,Normal} + F_{1,OSA}}{2}$$

가 본 경진대회의 평가 기준이 됩니다.

- $F_1$  Score 는 클래스별 Precision 과 Recall 의 조화 평균으로 구해지며, Precision 및 Recall 의 정의는 아래와 같습니다( $k$ 는 클래스, Normal 또는 OSA):

$$Precision_k = \frac{TP_k}{TP_k + FP_k}, Recall_k = \frac{TP_k}{TP_k + FN_k}$$

		PREDICTED classification				
		Classes	a	b	c	d
ACTUAL classification	a	TN	FP	TN	TN	
	b	FN	TP	FN	FN	
	c	TN	FP	TN	TN	
	d	TN	FP	TN	TN	

- 데이터별 추론 결과는 CSV 형식으로 제출하게 됩니다. 첫번째 열에 Test 용 데이터의 시리얼 번호가 입력되어 있는 CSV 파일 양식이 제공될 예정이며(아래의 좌측 이미지 참고), Normal 로 예측되는 데이터는 두번째 열의 값을 "Normal"로, OSA 리스크가 있는 것으로 예측되는 데이터는 두번째 열의 값을 "OSA"로 채워 제출해주시면 됩니다.

	A	B
1	Test_001	
2	Test_002	
3	Test_003	
4	Test_004	
5	Test_005	
6	Test_006	
7	Test_007	
8	Test_008	
9	Test_009	
10	Test_010	

	A	B
1	Test_001	OSA
2	Test_002	Normal
3	Test_003	Normal
4	Test_004	Normal
5	Test_005	OSA
6	Test_006	OSA
7	Test_007	OSA
8	Test_008	Normal
9	Test_009	OSA
10	Test_010	Normal

제공되는 CSV 파일 양식(좌), 답안 제출 예시(우, 실제 답안과는 무관함)

- 리더보드에는 가장 최근 제출의 점수가 반영됩니다.

- 정성평가
  - 순위권 내 팀들에 한해 발표 평가 (12/1 예정)
- 진행 유의사항
  - 본 대회는 학습을 위한 GPU 서버를 제공하며, 주어진 자원 내에서 결과물을 제출하여야 함 (외부 자원 활용 불가)
  - 의료 데이터 특성상 데이터는 다운로드가 불가능함
  - 현재 준비되어 있는 GPU 서버 12 대에 한정하여 대회를 진행할 것이며, 참가팀 12 팀 초과시, 서류심사를 통하여, 선별된 팀들에 한해 대회를 진행할 예정
  - 회원가입 뿐 아니라 참가 서류 작성(구글폼)을 완료한 참가자에 한하여 본선 대회를 진행할 수 있음 (미작성 혹은 허위 작성 적발시, 실격처리)

## F. 게시판 사용규칙 안내

- 질문은 이메일이 아닌, 대회 페이지 내 게시판을 통해서만 부탁드립니다. 게시판 사용시 정보 공유 차원에서 '개인정보 포함 문의 외에는 비밀글 사용이 제한'되며, 민감정보가 없는 질문인 경우 주최측에 의해 비밀글 기능이 해제되어 정보가 공유될 수 있습니다.

✧ 민감한 정보가 없다면 비밀글 대신, 공개글로 꼭 부탁드립니다!

- 게시판 질문시 제목에 질문 분류를 위한 [말머리] 를 꼭 달아주시길 부탁드립니다. 게시판 질문에 대응하는 여러 팀이 준비 중이기 때문에, 효율적인 대응을 위해 아래 5 가지 중 하나로 달아주시면 됩니다.
  - [임상] : 데이터의 의미에 관한 문의, Task 에 대한 문의, 전처리에 관한 문의 등
  - [플랫폼] : 주피터 노트북 오류, 코딩과정중에 발생하는 오류 등 서버 전반
  - [홈페이지] : MAIC 홈페이지 자체에서 발생하는 오류 (회원가입/정보 수정, 페이지 내 기능 등)
  - [운영] : 대회 규정, 팀 관련 이슈 등
  - [기타] : 위 사항 중 어느 곳에도 해당하지 않는다고 판단한 사항

☆ [말머리] 없는 질문은 답변이 없어도 책임 없음을 고지드립니다.

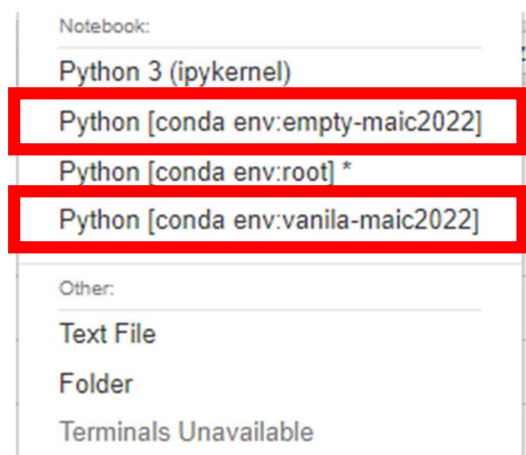
### 3. 환경 구성 설명

#### A. 기본 제공 사양

- CPU : 10Core
- RAM : 100Gb
- GPU : NVIDIA RTX A6000 GDDR6 ECC 48GB 1 개
- Storage : 5TB 까지
- 네트워크/디바이스 I/O 는 네트워크 환경에 따라 변경  
가능성이 있습니다

#### B. 기본 Anaconda 가상환경 제공

- vanilla-maic2022
  - Ubuntu 20.04
  - Cuda 11.7
  - Python 3.8.10
  - Pytorch 1.13.
  - Tensorflow 2.4.1
- empty-maic2022
  - python 3.8.10



- Cuda 11.7
- 라이브러리의 버전을 선택해서 설치 후 사용하고 싶은 분들은 해당 환경을 사용하시기 바랍니다.
- 대회 보안을 위해 터미널 환경은 제공하지 않습니다.

✧ 대회환경을 체크하기 위한 버전 및 라이브러리 체크 코드가 dataset 폴더에 제공됩니다