



user



내 PC



네트워크



휴지통



Chrome



메모장

# DEEP LEARNING

# DEE

# LEARN

# ING



# BEST EUNJIN



검색하려면 여기에 입력하십시오



-4°C

맑음



오후 5:57

2022-12-24





user



내 PC



네트워크



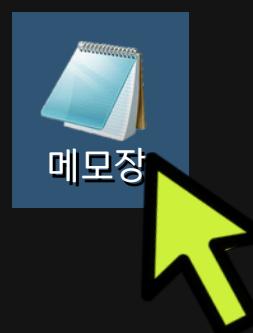
휴지통



Chrome



메모장





user



내 PC



네트워크



휴지통



Chrome



메모장

## 제목 없음 - Windows 메모장

- □ ×

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

< >

Ln 1, Col 1

100%

Windows (CRLF)

UTF-8

...



검색하려면 여기에 입력하십시오



-4°C

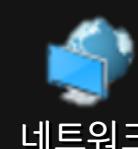
맑음



오후 5:57

2022-12-24





제목 없음 - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

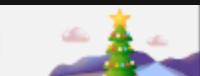
프로젝트\_가나'다나'마바사  
손글씨 인식 AI 모델 개발

(주)최고은진  
최은진, 최해정, 이종은, 이선영, 이지수, 김다나

Ln 1, Col 1 | 100% | Windows (CRLF) | UTF-8



검색하려면 여기에 입력하십시오



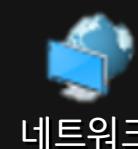
-4°C

맑음



오후 5:57  
2022-12-24





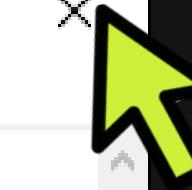
제목 없음 - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

프로젝트\_가나'다나'마바사  
손글씨 인식 AI 모델 개발

(주)최고은진  
최은진, 최해정, 이종은, 이선영, 이지수, 김다나

Ln 1, Col 1 | 100% | Windows (CRLF) | UTF-8



검색하려면 여기에 입력하십시오



-4°C

맑음



오후 5:57

2022-12-24





user



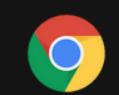
내 PC



네트워크



휴지통



Chrome



메모장

# DEEP LEARNING

# BEST EUNJIN



검색하려면 여기에 입력하십시오





user



내 PC



네트워크



휴지통



Chrome



메모장



손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 설정 가능

```
def 목차(contents):
    return 목차
목차(contents)
```

1. 팀 소개

2. 주제 선정 배경

3. 데이터 소개

4. 모델 소개

5. 학습 진행 과정

6. 학습 결과

7. 한계점 및 보완점

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 팀 소개

```
graph TD; A([대표이사  
최은진]) --- B([상무  
이선영]); B --- C([SW개발부]); C --- D([부장  
이지수]); C --- E([과장  
김다나]); C --- F([대리  
이종은]); C --- G([사원  
최해정])
```

대표이사  
최은진

상무  
이선영

SW개발부

부장  
이지수

과장  
김다나

대리  
이종은

사원  
최해정

0초 오후 11:44에 완료됨

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 설정 가능

# 주제 선정 배경

## 손글씨 인식 AI 모델 개발

많은 분야에서 문서가 전산화되고 있지만 아직까지 자필로 제출받아야 하는 문서가 존재하기 때문에 해당 문서들의 손글씨 전산화 및 데이터베이스에 저장함으로써 업무의 효율을 올릴 필요가 있음

산업 및 실생활에서 사용되는 다양한 한글 손글씨를 디지털 텍스트로 자동 확보하여, 사람이 반복적으로 처리해야하는 단순업무를 로봇 소프트웨어로 자동화 할 수 있는 손글씨 AI 모델 구축

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 설정 가능

# 데이터 소개

AI Hub

현대 한글 II, 172자를 사용한 폰트 50종의 글자체와 성별, 연령층별로  
직접 제작한 이미지 파일 구축 및 간판, 상표, 교통 표지판 등의 한글이 포함된 이미지 10만 장

#손글씨 # 현대 한글 # 인쇄체 # 표지판 # 상표 # OCR 기술 개발 # 이미지셋 # TTS 기술 개발

한국어 글자체 이미지

분야 영상이미지 유형 이미지

갱신년월: 2022-10 구축년도: 2019 조회수: 2,367 다운로드: 7,033 용량: 210.01 GB

검색하려면 여기에 입력하십시오

0초 오후 11:44에 완료됨

-8°C 대체로 흐림 오후 11:47 2022-12-24



손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 데이터 소개



손글씨

인쇄체

Text in the Wild

[문장]	train	= 42,240	train	= 34,258	image	= 60,000
	validation	= 5,280	validation	= 4,030	train	= 139,109
	test	= 5,280	test	= 2,015	validation	= 29,813

---

[단어]	train	= 71,999	train	= 81,452	test	= 27,678
	validation	= 17,999	validation	= 21,020		
	test	= 6,296	test	= 21,020		

검색하려면 여기에 입력하십시오

0초 오후 11:44에 완료됨

-8°C 대체로 흐림

오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 설정 가능

# 데이터 소개\_json 형태

[손글씨/인쇄체]

[문장]

```
{'id': '00000044',  
 'image_id': '00000044',  
 'text': "의할 예정이라고 8일 밝혔다. 이 의원이 밝힌 개정안 내용은 '일제하',  
 'attributes': {'type': '문장', 'gender': '여', 'age': '28', 'job': '직장인'}},
```

[단어]

```
{'id': '00013854',  
 'image_id': '00013854',  
 'text': '말하다',  
 'attributes':  
 {'type': '단어(어절)',  
 'gender': '남', 'age': '24', 'job': '대학생'}}},
```

[text in wild]

```
{'id': '00000001',  
 'width': 1920,  
 'height': 1440,  
 'file_name':  
 'FFF2F34A08347075F55E72B240EFE691.jpg',  
 'type': 'book'}
```

```
{'id': '00000001',  
 'image_id': '00000001',  
 'text': '냉',  
 'attributes': {'class': 'character'},  
 'bbox': [713, 272, 76, 92]}}
```

검색하려면 여기에 입력하십시오

0초 오후 11:44에 완료됨

-8°C 대체로 흐림

오후 11:47  
2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 모델 소개

[OCR(Optical Character Recognition)]

사람이 직접 입력하지 않고 스캐너를 통해 이미지 형태로 읽어들여 데이터의 내용을 분석하고 그림 영역과 글자 영역으로 구분한 후 글자 영역의 문자들을 일반 문서 편집기에 수정, 편집이 가능한 텍스트의 형태로 변환하여 주는 자동 입력 시스템

스캔된 인쇄 문서 내 text 식별 - 인쇄체(printed)와 필기체(handwritten), 일상에서 캡쳐된 이미지에서 text를 식별 - text in the wild

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-Ixo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 설정 가능

# 모델 소개 [OCR 과정]

```
graph LR; A((전처리)) --> B((문자 검출)); B --> C((문자 인식))
```

전처리(pre-processing)

- 글자들이 잘 보여질 수 있게 밝기나 색과 같은 영상의 메타데이터를 변화시킴

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림

오후 11:47  
2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 설정 가능

# 모델 소개 [OCR 과정]

```
graph LR; A[전처리] --> B[문자 검출]; B --> C[문자 인식]
```

문자 검출(Text Detection)

- 글자들이 존재하는 위치를 찾아내 bounding box로 묶기 → 인식 모델의 입력 데이터
- CNN(Convolutional Neural Network)계열의 모델을 주로 사용

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 모델 소개 [OCR 과정]

```
graph LR; A[전처리] --> B[문자 검출]; B --> C[문자 인식];
```

문자 인식(Text Recognition)

- bounding box 안에 글자 추출
- RNN(Recurrent Neural Network)계열의 모델 주로 사용

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-Ixo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 모델 소개 [OCR 과정]

```
graph LR; A[전처리] --- B[문자 검출]; B --- C[문자 인식]; C --> D["문자 인식(Text Recognition)"]
```

문자 인식(Text Recognition)

deep-text-recognition-benchmark

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 진행 과정 [데이터 전처리]

00012867.png



train\_labels.json

```
{'id': '00012867',  
 'image_id': '00012867',  
 'text': '갈등',  
 'attributes':  
 {'type': '단어(어절)',  
 'gender': '남', 'age': '46', 'job': '회사원'}}
```

labels\_train.txt

00012867.png 갈등

LMDB 메모리 맵핑 데이터베이스는 키-값 저장소 형태로 포함된 트랜잭션 데이터베이스를 제공하는 소프트웨어 라이브러리

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과

[손글씨]

원본이미지

뒤진다 강한하

총 30,000번 학습 결과 : 정확도 97.052

```
[10000/10000] Train loss: 0.08515, Valid loss: 0.15842, Elapsed_time: 3066.72009  
Current_accuracy : 92.621, Current_norm_ED : 0.97  
Best_accuracy : 92.621, Best_norm_ED : 0.97
```

---

```
[20000/20000] Train loss: 0.00550, Valid loss: 0.09038, Elapsed_time: 6077.60727  
Current_accuracy : 96.467, Current_norm_ED : 0.98  
Best_accuracy : 97.052, Best_norm_ED : 0.99
```

---

Ground Truth	Prediction	Confidence Score & T/F
높아지다	높아지다	0.9020 True
허가	숙가	0.9082 False
엄청나다	엄청나다	0.9965 True
광고	광고	0.9997 True
대출	대출	0.9998 True

---

# Word

[인쇄체]

원본이미지

비행 주위

손글씨 단어 학습 모델에 인쇄체 단어 전이 학습

18,000번 학습 결과 : 정확도 99.981

```
[16000/16000] Train loss: 0.00005, Valid loss: 0.00145, Elapsed_time: 3431.75488  
Current_accuracy : 99.962, Current_norm_ED : 1.00  
Best_accuracy : 99.981, Best_norm_ED : 1.00
```

---

Ground Truth	Prediction	Confidence Score & T/F
더	더	1.0000 True
항의	항의	0.9999 True
개	개	1.0000 True
마루	마루	1.0000 True
활인	활인	0.9999 True

---

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림

오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과

## Word

### test data로 테스트 진행 결과

#### [손글씨]

	Labels	pred	ox	confidence_score
0	갑	갑	True	0.999900
1	강화하다	강라하다	False	0.397900
2	그이	그이	True	0.999600
3	수십	수십	True	0.999600
4	수집하다	수집하다	True	0.904400
5	보전	보전	True	0.998800
6	부드럽다	부드럽다	True	0.753600
7	나오다	나오다	True	0.379100
8	둘	둘	True	0.991000
9	뒤지다	뒤지다	True	0.990500
10	예상하다	예상하다	True	0.998100

#### 정확도 : 80.627

#### [인쇄체]

	Labels	pred	ox	confidence_score
11	예선	예선	True	0.999800
12	오가다	오가다	True	0.999800
13	전개하다	전개하다	True	0.997700
14	망하다	망하다	True	0.991500
15	후보	후보	True	0.999800
16	흔적	흔적	True	0.863700
17	버려지다	내려지다	False	0.758400
18	닫다	와	False	0.836900
19	얼다	엄다	False	0.884900
20	여고생	여고생	True	0.996500

#### 정확도 : 99.981

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 설정 가능

# 학습 결과

## Word

### 실제 데이터를 형성하여 최종 테스트 결과

	img	id	label	conf
0			가	0.9990
1			자	0.9999
2			원	1.0000
3			궁	0.4027
4			쪽	0.1835
5			술	0.4667
6			문	0.8842
7			안	0.6382
8			새	0.9995
9			해	0.4539
10			존재	0.9932

검색하려면 여기에 입력하십시오

0초 오후 11:44에 완료됨

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과

## Word

### 실제 데이터를 형성하여 최종 테스트 결과

	img	id	label	conf
0			가	0.9990
6			문	0.8842
4			금	0.1835
9			새	0.4539
5			몽	0.4667
10			존재하다	0.9932

제공된 데이터 이외의 데이터에 대해서는 정답률이 높지 않음

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-Ixo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과

## Sentence

[인쇄체]

### 원본이미지

겠습니다. 충남과 남부에 20에서 60mm, 수도권과 강원, 충북에는 5

인쇄체 문장 25,000번 전이 학습 정확도 99.132

```
[5000/5000] Train loss: 0.00019, Valid loss: 0.00150, Elapsed_time: 2709.12496
Current_accuracy : 99.132, Current_norm_ED : 1.00
Best_accuracy : 99.132, Best_norm_ED : 1.00
```

Ground Truth	Prediction	Confidence Score & T/F
나있고미세먼지농도도보통단계입니다하지만전남과제주지역	나있고미세먼지농도도보통단계입니다하지만전남과제주지역	0.4212 True
시키기위해위장전입한것은사실이나오씨의도움을받아b고교에	시키기위해위장전입한것은사실이나오씨의도움을받아b고교에	0.3812 True
장인이목희열린우리당의원은국회절차법에따라법안을심사할	장인이목희열린우리당의원은국회절차법에따라법안을심사할	0.6390 True
습니다바다의물결은동해중부먼바다에서최고4m까지매우높게	습니다바다의물결은동해중부먼바다에서최고4m까지매우높게	0.1166 True
대북정책을믿고국론을하나로모아달라고당부습니다kbs뉴스	대북정책을믿고국론을하나로모아달라고당부습니다kbs뉴스	0.3703 True

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과

## Sentence

[손글씨]

### 원본이미지

점을 예견했다. 이처럼 물관리에 심각한 문제 가 생겼을 때에도 수자원공

### 손글씨 문장 15,000번 전이 학습 정확도 81.553

```
[10000/10000] Train loss: 0.01125, Valid loss: 0.05307, Elapsed_time: 11358.67981  
Current_accuracy : 81.326, Current_norm_ED : 0.99  
Best_accuracy : 81.553, Best_norm_ED : 0.99
```

Ground Truth	Prediction	Confidence Score & T/F
인하대는의예과남학생들의여학생집단성희과관련해9일공식	인하대는의예과남학생들의여학생집단성희과관련해9일공식	0.2335 True
흡수된다리스크검사지원국은폐지된다금감원은또소비자서비스	흡수된다리스크검사지원국은폐지된다금감원은또소비자서비스	0.1210 True
국에서출품한190편의영화가상영된다5월5일부터는전주한지문	국에서출품한190편의영화가상영된다5월5일부터는전주한지문	0.3267 True
는이번방은안증근의사관련레포트우수대학생13명전국중	는이번방은안증근의사관련레포트우수대학생13명전국중	0.1275 True
제하는것은표현의자유가로막을뿐이며정부가반정부시위때	제하는것은표현의자유가로막을뿐이며정부가반정부시위때	0.1288 False

0초 오후 11:44에 완료됨

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과

## Sentence

[인쇄체] 테스트 정확도 : 81.7911

### test data로 테스트 진행 결과

Labels	pred	ox	confidence_score
가을막이붕괴사고는수차례안전보강공사가필요하다는지적을무	가을막이붕괴사고는수차례안전보강공사가필요하다는지적을무	True	0.376100
로보여개막식행사에참여하는분들은추위에대비잘해주셔야	로보여개막식행사에참여하는분들은추위에대비잘해주셔야	True	0.259000
드러다한편c군의아버지는아들을강등구소재고교에전입학	드러다한편c군의아버지는아들을강등구소재고교에전입학	True	0.217200
부모사이에친분관계가있을가능성을시사다또b고교의한	부모사이에친분관계가있을가능성을시세다또b고교의인	False	0.084600
다	다	True	0.999800

### 실제 데이터를 형성하여 최종 테스트 결과(img / label / conf)

최은진 최해정 이종은 이선영 이지수 김다나 최은전피해정이송은이선명이지수강다나 0.0433

전장연, 내년 1월부터 시위 재개한다... 출근길 지하철 대 전장면내년월부터시위제기한다들은길지하결대 0.0056

채태인 김민성 김상수 사인 앤드 트레이드, 키움은 이유가 있었다... 제대인김반성김사수사인드트레이드키옹은이유가있다 0.0171

여당 전당대회 선관위원장에 유흥수 상임고문 예당전당대회선광위장에유등수생업고문 0.0016

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과 Sentence

## [손글씨] 테스트 정확도 : 81.6321

### test data로 테스트 진행 결과

Labels	pred	ox	confidence_score
대변인을맡아투쟁을이끌다김위원장은7일내일신문과통화에	대변인을맡아투쟁을이끌다김위원장은7일내일신문과통화에	True	0.336800
리고모레밤에는경기북부와강원영서북부에비가내리습니다	리고모레밤에는경기북부와강원영서북부에비가내리습니다	True	0.600300
경영실적보고서평가편람내용과다르게작성하지말라고주의	경영실적보고서평가편람내용과다르게작성하지말라고주의	True	0.187400
하기때문에꺼리는것으로풀이다노동사무소관계자는근로시간	하기때문에꺼리는것으로풀이다노동사무소관계자는근로시간	True	0.265200
무가분리되면서은행보험금투자중소서민등각권역별로감독	무가분리되면서은행보험금투자중소서민등각권역별로감독	True	0.298500

### 실제 데이터를 형성하여 최종 테스트 결과(img / label / conf)

미리저리 줄려 온 세상을 훑 뒤집어 우리 엄마 말도 잘 안 들던 내가	이크지풀러모씨상을충뒤검어우리영다필도잘만들면내가	0.0000
침이다한편금감원은 이에앞서새로운임원진을 사실상 확정다	참에대한면금감원은이에앞서로운임원진을사실상확정다	0.0007
침이다 한편 금감원은 이에 앞서 새로운 임원진을 사실상 확정다.	침이다한편금감원은이에앞서새로운임원진을사실상확정다	0.2913
한 강찰조사 진행하고 있지만 이번 사표가 수리되면 강찰은 정지된	한강찰조사진행하고있지만이번사표가수리되면강찰은정지된	0.1257

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과

Sentence

[손글씨] 테스트 정확도 : 81.6321

test data로 테스트 진행 결과

데이터 수가 적어 과적합 발생  
학습된 문장이 아닌 경우 예측 어려움

## 실제 데이터를 형성하여 최종 테스트 결과(img / label / conf)

미리저리 줄려 온 세상을 훑 뒤집어 우리 엄마 말도 잘 안 들던 내가

이크지풀러모씨상을충뒤검어우리영다필도잘만들면내가 0.0000

침이다한편금감원은 이에앞서새로운임원진을 사실상 확정다

참에대한면금감원은이에앞서로운임원진을 사실상 확정다 0.0007

침이다 한편 금감원은 이에 앞서 새로운 임원진을 사실상 확정다.

침이다한편금감원은이에앞서새로운임원진을 사실상 확정다 0.2913

한 감찰조사 진행하고 있지만 이번 사표가 수리되면 감찰은 정지된

한감찰조사진행하고있지만이번사표가수리되면감찰은정지된 0.1257

✓ 0초 오후 11:44에 완료됨



검색하려면 여기에 입력하십시오



-8°C 대체로 흐림

오후 11:47  
2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과

## 원본이미지

## 크롭이미지

# Text In The Wild

### 학습 결과 : 정확도 84.210

```
[1/15000] Train loss: 118.64859, Valid loss: 113.50572, Elapsed_time: 8.51595  
Current_accuracy : 0.000, Current_norm_ED : 0.00  
Best_accuracy : 0.000, Best_norm_ED : 0.00
```

Ground Truth	Prediction	Confidence Score & T/F
전주포차	셋체f체새f새땀f땀새땀목없목땀f덟프f흉압	0.0000 False
nh	셋범셋새없흰없목없팬머꼬권	0.0000 False
치맛속까지	셋폐없땀없흉탕	0.0000 False
팔찌	셋폐없맨흉권	0.0000 False
열쇠박사	셋딴없쇼없싸압	0.0000 False

```
[15000/15000] Train loss: 0.19837, Valid loss: 0.53552, Elapsed_time: 8682.35031  
Current_accuracy : 84.210, Current_norm_ED : 0.91  
Best_accuracy : 84.210, Best_norm_ED : 0.91
```

Ground Truth	Prediction	Confidence Score & T/F
종로3가	종로3가	0.7702 True
소화	소화	0.9906 True
삼겹살	삼겹살	0.8983 True
제한높이	제한높이	0.9664 True
향	향	0.9911 True

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 학습 결과

전이 학습 결과  
처음부터 정확도 83.880 → 86.660

```
[1/15000] Train loss: 0.42562, Valid loss: 0.58731, Elapsed_time: 8.52964
Current_accuracy : 83.880, Current_norm_ED : 0.91
Best_accuracy : 83.880, Best_norm_ED : 0.91
-----
Ground Truth | Prediction | Confidence Score & T/F
-----
start | start | 0.6930 True | 0.5943 True
컴퓨터활용능력 | 컴퓨터활용능력 | 0.9390 True
배우기 | 배우기 | 0.9975 True
옮김 | 옮김 | 0.9976 True
호텔 | 호텔 |
```

```
[15000/15000] Train loss: 0.00891, Valid loss: 0.65631, Elapsed_time: 7856.46619
Current_accuracy : 86.365, Current_norm_ED : 0.92
Best_accuracy : 86.660, Best_norm_ED : 0.92
-----
Ground Truth | Prediction | Confidence Score & T/F
-----
옮김 | 옮김 | 0.9997 True
greatest | greatest | 0.5321 True | 0.5142 True
세라마이드 | 세라마이드 | 0.9630 True
버스외 | 버스외 | 0.8437 True
아론적 | 아론적 |
```

end the training

CPU times: user 34.1 s, sys: 5.24 s, total: 39.3 s

Wall time: 2h 11min 51s

# Text In The Wild

테스트 결과  
정확도 : 87.367

	Labels	pred	O/X	confidence_score
0	이청준	이경구	False	0.224200
1	전집	전집	True	0.792400
2	당신들의	당신들의	True	0.988800
3	천국	천국	True	0.980300
4	문학과지성사	불학물지차	False	0.002100
5	프란츠	프란츠	True	0.999200
6	카프카	카프카	True	0.999000
7	소송	소송	True	0.999100
8	penguin	penguin	True	0.785700
9	따뜻함과	마랑과과	False	0.289500
10	말과	말과	True	0.997200
11	글에는	글에는	True	0.942500

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

댓글 공유 설정

RAM 디스크 설정 가능

# 학습 결과

## Text In The Wild

### 실제 이미지 테스트 결과

번호	원본 이미지	예측 텍스트	정확도
0		똑똑부동산	0.9778
1		똑똑이	0.9826
2		똑똑이와	0.8957
3		똑똑이세트	0.8032
4		opencampo2	0.8690
5		오픈캠프 02	0.4372
6		ACademy	0.5538
7		Software	0.2208
8		Seoul	0.8920
9		추락주의	0.5017
10		마스터	0.8468

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

RAM 디스크 설정 가능

# 정리 \_전이학습 결과

단어 문장 Text in the Wild

	단어	문장	Text in the Wild
before	97.052	99.132	84.210
after	99.981	81.553	86.660

전이학습 순서 손글씨 → 인쇄체 인쇄체 → 손글씨 데이터셋 추가학습

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 설정 가능

# 한계점 및 보완점

1. text-detection을 생략하고 text-recognition만 진행

2. 학습하지 못한 단어 및 문장 패턴인 경우 정확도가 떨어짐

미리저리  
줄려 온 세상을  
휙 넘침어  
우리 엄마 말도  
잘 안 들던 내가  
이크지풀러모씨상을충뒤검어우리영다필도잘만들면내가

금쪽이

꼼특이

3. 시간 및 비용적인 문제로 학습에 한계가 존재함

4. 학습횟수를 늘리는 것 외에 증강데이터 사용 등 다른 보완 방법 고민 필요

0초 오후 11:44에 완료됨

검색하려면 여기에 입력하십시오

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 설정 가능

print('지금까지 발표를 들어주셔서 감사합니다')

img	label	conf
지금까지 발표를 들어주셔서 감사합니다	대증까지 발표등록하기 주제증사 확대	0.0001
지금까지 발표를 들어주셔서 감사합니다	지금까지 발표들어주제증시합니다	0.0152
지금까지 발표를 들어주셔서 감사합니다	지금까지 발표등록해주시기 바랍니다	0.0137

검색하려면 여기에 입력하십시오

0초 오후 11:44에 완료됨

-8°C 대체로 흐림 오후 11:47 2022-12-24

손글씨 인식 AI 모델 개발.ipynb

colab.research.google.com/drive/11vG0lcWhq95D1-lxo0fgr1AiuRCGGCQG#scrollTo=TnJNenPnllPe

손글씨 인식 AI 모델 개발.ipynb ☆

파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨

+ 코드 + 텍스트

✓ RAM 디스크 | 수정 가능

print('지금까지 발표를 들어주셔서 감사합니다')

img	label	conf
지금까지 발표를 들어주셔서 감사합니다	대증까지 발표등록하기 주제증사 확대	0.0001
지금까지 발표를 들어주셔서 감사합니다	지금까지 발표들어주제증시합니다	0.0152
지금까지 발표를 들어주셔서 감사합니다	지금까지 발표등록해주시기 바랍니다	0.0137

검색하려면 여기에 입력하십시오

0초 오후 11:44에 완료됨

-8°C 대체로 흐림 오후 11:47 2022-12-24



user



내 PC



네트워크



휴지통



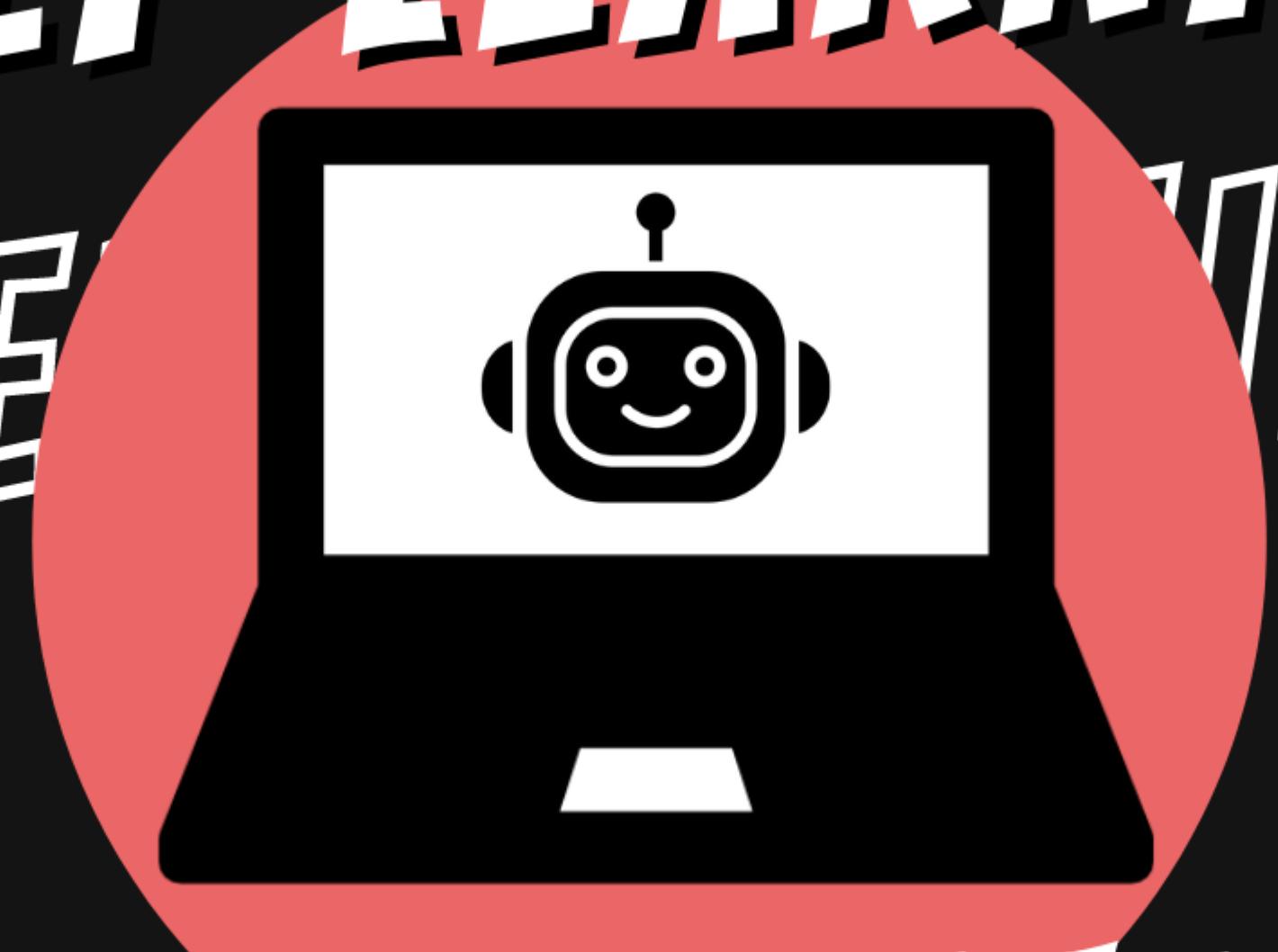
Chrome



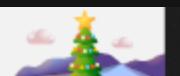
메모장

# DEEP LEARNING

# BEST EUNJIN



검색하려면 여기에 입력하십시오



-4°C

맑음



오후 5:57  
2022-12-24





user

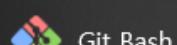


내 PC

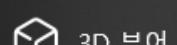


네트워크

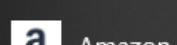
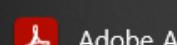
최근에 추가한 앱



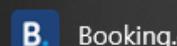
#



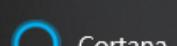
A



B



C

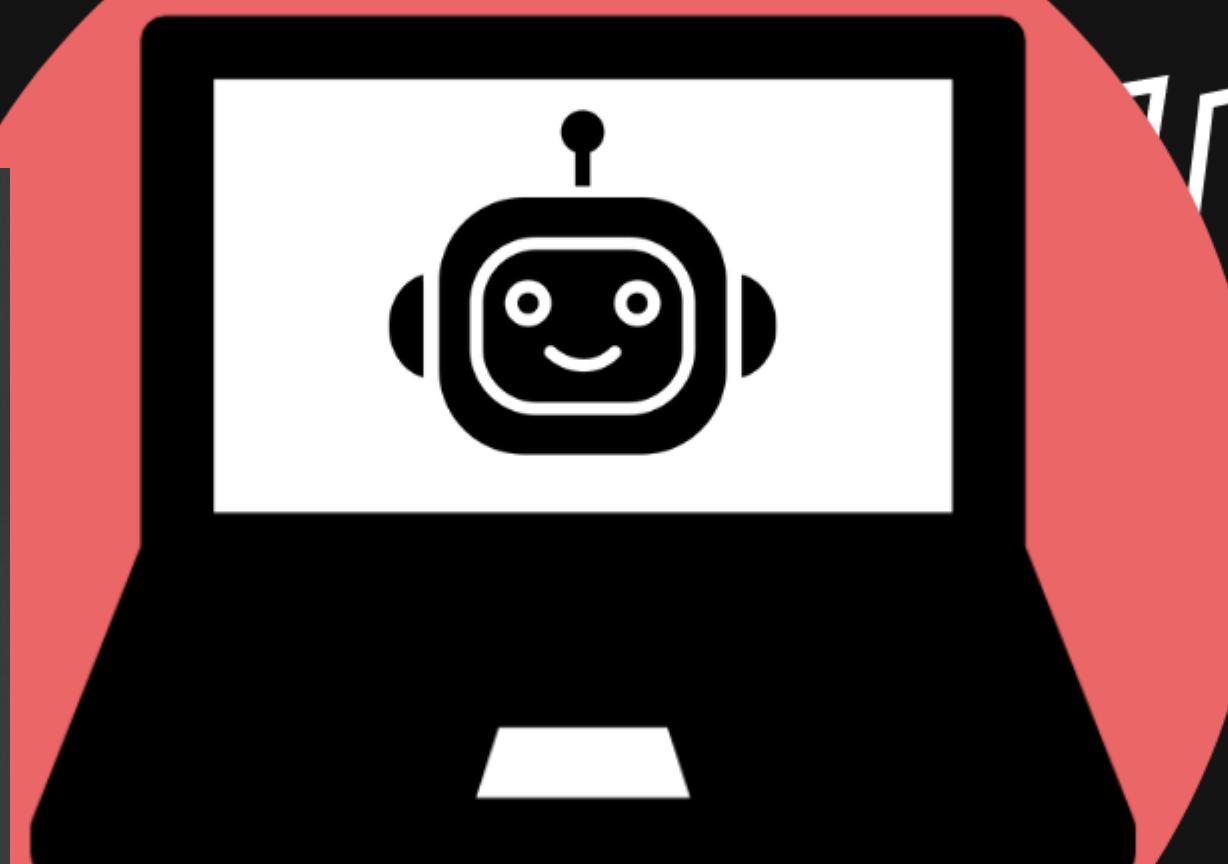


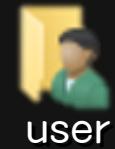
E



# DEEP LEARNING

# T EUNJIN





user



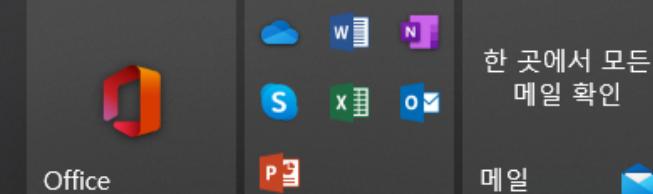
내 PC



네트워크

☰ 시작

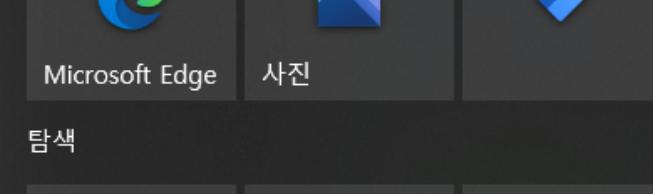
생산성



한 곳에서 모든  
메일 확인

메일

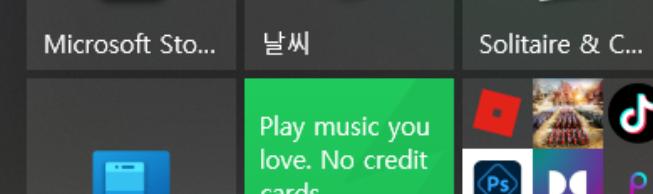
탐색



날씨

Solitaire & C...

▼

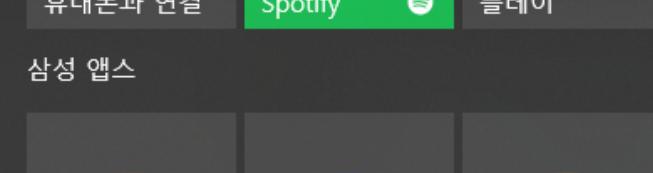


Microsoft Sto...

날씨

Solitaire & C...

▼



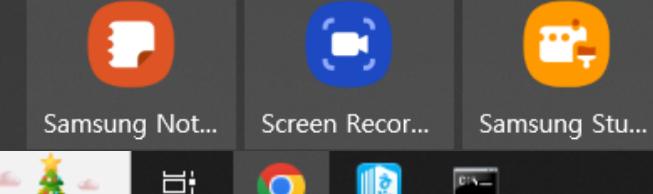
Samsung Not...

Screen Recor...

Samsung St...

삼성 앱스

▼



Samsung Not...

Screen Recor...

Samsung St...

☰ 설정

☰ 전원

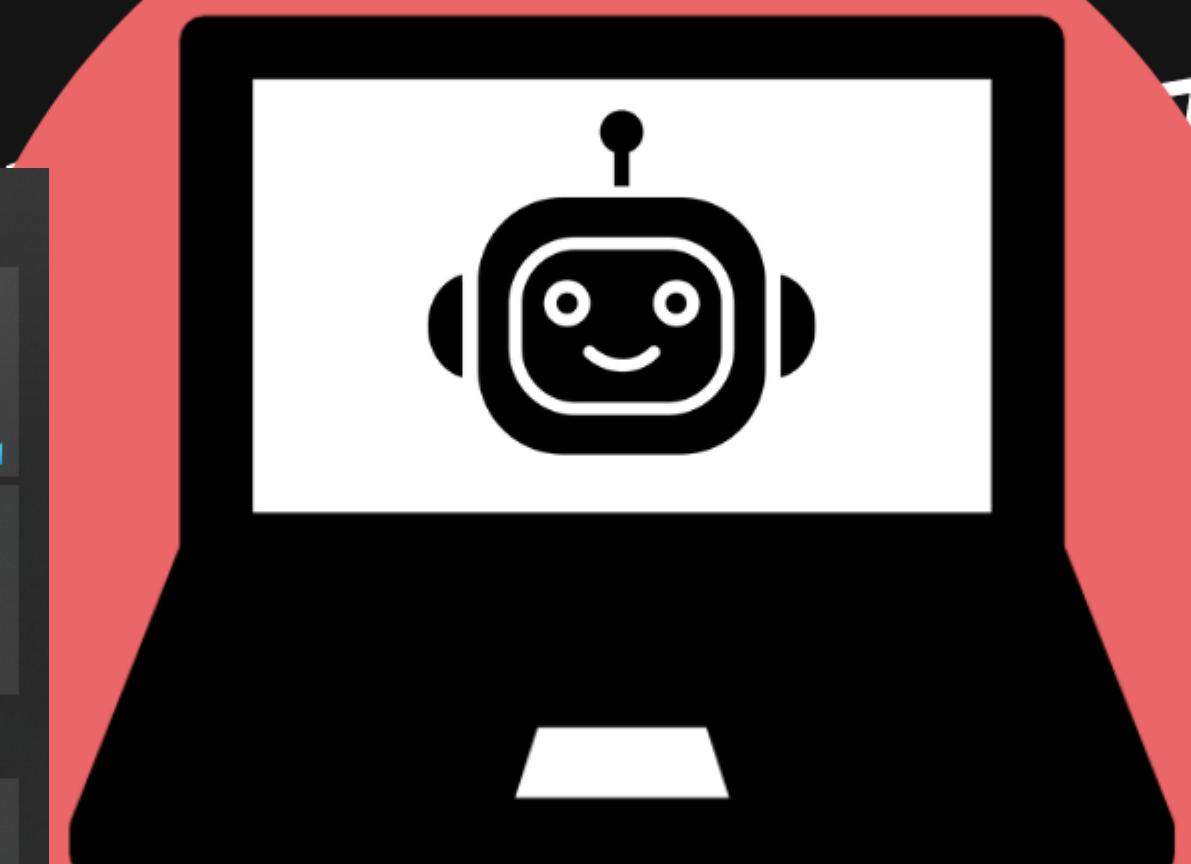


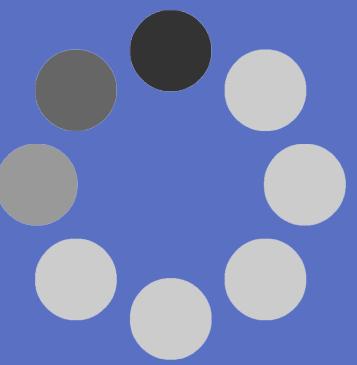
검색하려면 여기에 입력하십시오



# DEEP LEARNING

# T EUNJIN





종료 중

