

# 주진호

*데이터로 사람과 문제를 잇는 연구자를 지향합니다.*

CONTACT

---

EMAIL

wlsgh20728@naver.com

Web Site & Blog

<https://wlsghdh.github.io>

<https://edu-data.tistory.com>

PHONE

010-4919-1534

GITHUB

github.com/Wlsghdh

ADDRESS

seongnam



# 주진호

데이터과학부 · 학점 4.34 / 4.5

3

PUBLICATIONS

4

PROJECTS

5

AWARDS

4

EXPERIENCES

## Publications

Why Deep ResNets Train Successfully: Self-Selection of Effective Depth Enabled by Skip Connections

소규모 산업 결함 검출을 위한 생성형 AI 증강 효과

통합 전처리 파이프라인을 활용한 멀티모달 가스 센서 데이터셋에서의 모델 성능 비교

## Awards

2025.12	한국정보과학회 장려상	정보과학회
2025.08	신약 개발 20위 (상위 4%)	데이콘
2025.10	주식차트 데이터 수집_1 1등	수원대학교 DSML
2025.11	주식차트 데이터 수집_2 1등	수원대학교 DSML
2024.08	AI/Develops 우수상	수원대학교 DSML

## SKILLS

	<div style="width: 30%;"></div>	PR, Branch, Commit, Issue ..
	<div style="width: 70%;"></div>	Pytorch, LLM, Sciki-learn, Numpy Rag, Selenium, Docker ..
	<div style="width: 40%;"></div>	Linux Command
	<div style="width: 60%;"></div>	Hugging face, YOLO, detectro2 mmdetection, CV

## EDUCATION

2021.03 - 2027.02	수원대학교 데이터과학부 학사
2019.03 - 2021.02	성일 고등학교
Academic Excellence Scholarship, [Suwon University]	

Fall 2021, Fall 2024, Spring 2025, Fall 2025, Spring 2026  
Awarded for outstanding academic performance (5 semesters)

2026

Arxiv, BMVC accept 대기중

**Why Deep ResNets Train Successfully: Self-Selection of Effective Depth Enabled by Skip Connections**

주진호, 안홍렬

[https://github.com/Wlsghdh/Learnable\\_Residual\\_Scaling](https://github.com/Wlsghdh/Learnable_Residual_Scaling)

제1저자

---

2026

한국정보과학회

**소규모 산업 결함 검출을 위한 생성형 AI 증강 효과**

주진호, 임대윤, 양진우, 안홍렬

<https://github.com/Wlsghdh/VISION-Instance-Seg>

제1저자

---

2025

한국정보과학회

**통합 전처리 파이프라인을 활용한 멀티모달 가스 센서 데이터셋에서의 모델 성능 비교**

맹영민, 주진호, 윤재훈, 정우창, 안홍렬

<https://github.com/Ahn-Laboratory/Gas-Leakage-Detection>

제2저자

---

# Why Deep ResNets Train Successfully: Self-Selection of Effective Depth Enabled by Skip Connections

CIFAR-10/100 ImageNet ResNet-50/101/152/200

## 1. 문제인식 - 이해결 질문

“같은 깊이의 일반 신경망은 처참히 실패하는데, 왜 ResNet은 200층을 쌓아도 퇴화하지 않는가?”

표준 설명 “스킵 연결이 그래디언트 흐름을 안정화한다”는 기술적으로는 옳지만 메커니즘적으로 불완전하다. **스킵 연결이 단지 그래디언트를 흘려보낸다면, 깊은 ResNet은 그 깊이로 무엇을 하고 있는가?**

기존 잔차 스케일링 (ReZero · Skiplnit · Fixup · LayerScale) 은  $F(x)$ 만 스케일링하고 항등 부분은 가중치 1로 고정 → 블록별 사용량을 직접 읽어낼 수 없음.

## 2. 핵심 명제 & LRS – “측정 도구”

Deep ResNet은 **명목 깊이(L)보다 훨씬 작은 효과적 깊이를 자율 선택**하기 때문에 학습에 성공하며, 스킵 연결은 그 자율 선택을 가능케 하는 구조적 메커니즘이다.

$$y = \alpha \cdot F(x) + (1 - \alpha) \cdot x, \quad \alpha = \sigma(\theta) \in (0, 1)$$

블록당 단일 학습 가능 스칼라  $\theta$  · 두 계수 합이 1  $\Rightarrow \alpha_i$  = 블록  $i$ 의 사용량 그 자체

기존 방법과 달리 LRS는  $F(x)$ 와  $x$ 를 동시에 조정하는 블록 결합. 사후 분석 불필요 **학습된  $\alpha$ 가 곧 답이다.**

## 3. 증거 A - Plain은 자율 선택을 못해서 실패. (Optimal depth 찾지 못 한다.)

CIFAR-100 / 200층, 초기화 시점 블록별 그래디언트 L2 Norm

Plain (스킵 없음)

**$3.5 \times 10^{12}$  배**

학습 불가능- 폭발

ResNet ( 표준 )

**673배**

스킵 연결이 분포 안정화

LRS (학습 후)

**$\alpha_{후기} \approx 0$**

신경망이 후기 블록을 자율 우회

스킵이 없는 일반 신경망은 깊은 블록과 얇은 블록의 그래디언트가 1조 배 차이 나서 학습 자체가 불가능. 스킵을 넣으면 673배 수준으로 안정화되지만, LRS 측정으로 관찰하면 신경망은 그 안정성을 “모든 블록 활용”이 아니라 “후기 블록 우회”에 사용한다는 새로운 사실이 드러남.

## Why Deep ResNets Train Successfully: Self-Selection of Effective Depth Enabled by Skip Connections

CIFAR-10/100 ImageNet ResNet-50/101/152/200

### 4. 증거 B - 깊이가 늘수록 $\alpha$ 는 단조 감소 (스케일링 법칙)

깊이 L	# 블록	CIFAR-100 $\bar{\alpha}$	ImageNet $\bar{\alpha}$	활성 블록 ( $\alpha > 0.3$ )	$D_{\text{eff}} = \Sigma \alpha_i$
50	16	0.342	0.303	6-7	5.4 (34%)
101	33	0.183	0.166	6	6.0 (18%)
152	50	0.13	—	5-6	6.7 (13%)
200	66	<b>0.11</b>	—	<b>5-6 / 66</b>	<b>7.5 (11%)</b>

블록 수 4배(16→66) 증가에도 효과적 깊이는 단 1.4배(5.4→7.5)만 증가. 200층 신경망은 사실상 5-6개 블록만 사용한다. ImageNet도 동일 추세 (d50→d101에서 CIFAR 1.87x vs ImageNet 1.83x) **데이터셋 가설을 배제하는 결정적 증거.**

### 5. 증거 (c) - " $\alpha$ 기반 가지치기"로 본질성 검증

임계값	제거 블록	Acc (미조정)	Acc (+조정)	FLOPs	지연
원본	0 / 66	80.54%	—	100%	100%
$\alpha < 0.03$	2 / 66	<b>80.54%</b>	<b>80.55% (+0.01)</b>	97%	98%
$\alpha < 0.08$	<b>37 / 66 (56%)</b>	3.97%	<b>75.91% (-4.6)</b>	<b>46%</b>	<b>49%</b>
$\alpha < 0.10$	48 / 66 (73%)	3.13%	74.74% (-5.8)	30%	32%

$\alpha < 0.03$  블록은 재학습 없이도 정확도 손실 0%로 제거 측정 가능성에 대한 직접적 의미 검증. 또한 항등 초기화 vs 무작위 초기화로 학습 시 수렴  $\bar{\alpha}$ 가 0.118 vs 0.114로 거의 동일 → 출발점 무관 = 본질적 현상.

## Why Deep ResNets Train Successfully: Self-Selection of Effective Depth Enabled by Skip Connections

CIFAR-10/100 ImageNet ResNet-50/101/152/200

### 6. 실용적 임팩트 분석 도구이자 가지치기 기준

1.  $\alpha$  기반 가지치기 = 학습 부산물이 곧 블록 중요도 기준 (재학습 없이도 일부 제거 가능).
2. 항등 선호 초기화 ( $\alpha_0 \approx 0.12$ ) 필수    잔차 선호 ( $\alpha_0 \approx 0.88$ )는 D200에서 학습 붕괴 ( $37.5 \pm 29.2\%$ ).
3. NAS / 효율 아키텍처 설계 가이드: 용량은 스테이지 내부가 아닌 스테이지 경계(다운샘플링)에 할당해야 함 (스테이지 경계 블록  $\alpha \in [0.42, 0.96]$  vs 내부  $\alpha \in [0.06, 0.15]$ ).

---

### 7. Degradation problem 재해석

Degradation problem 해결해야 할 실패가 아니다. 그것은 신경망이 불필요한 깊이의 사용을 올바르게 거부하는 현상이다. 깊은 신경망이 작동하는 이유는 대부분의 레이어를 사용하지 않도록 허용될 때다 사용되지 않는 용량은 낭비가 아니라 예비이며, 스킵 연결은 이를 실행 가능한 설계 선택으로 만드는 메커니즘이다.

## 소규모 산업 결함 검출을 위한 생성형 AI 증강 효과

### 1. 문제인식 - 산업 결함 검출의 구조적 데이터 부족

“클래스당 20장. 이 데이터로 어떻게 결함을 검출하는 모델을 학습시킬 것인가?”

실제 산업 현장에서 불량 발생 빈도 자체가 낮고, 픽셀 단위 라벨링은 시간·비용이 크다. 대표 공개 데이터셋 VISION-Datasets조차 클래스당 수십 장이 표준.

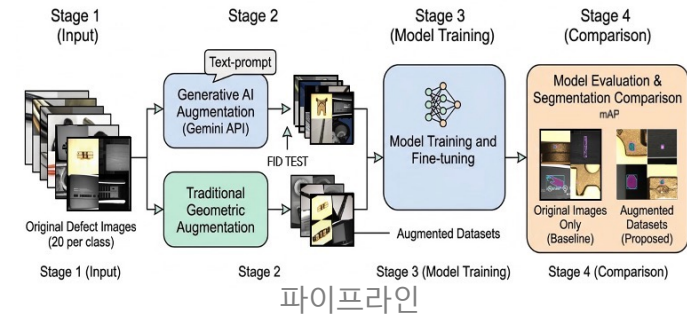
기존 해법 = 전통 증강 (회전·색상·반전). 단순·저비용이지만 만들어내는 이미지가 원본의 픽셀 변형에 머무름. 한정된 시각 패턴을 반복 학습 → 과적합 위험.

### 2. 핵심 가설 - “데이터의 출처가 양보다 중요하다”

같은 분량의 증강이라도, 원본의 시각 패턴 범위를 벗어나는 의미 수준 다양성이 소규모 결함 조건에서 결정적이다.

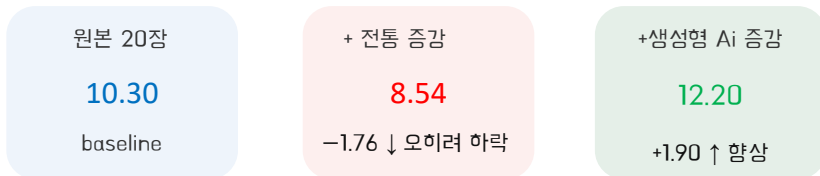
전통 증강 (Albumentations)  
회전·반전·색상 변환  
→ 원본의 픽셀 변형에 머무름

생성형 AI 증강 (Gemini 2.0)  
결함 제거 → 새 위치·모양으로 합성  
→ 원본에 없던 새 패턴 생성



### 3. 실험 1 단독 비교에서 정반대 방향 발견

Mask R-CNN · 클래스당 +125장 증강 · mAP (3개 클래스 평균)



같은 양의 데이터를 추가했는데 효과가 정반대. 3개 클래스(주조 개재물·콘솔 오염·목재 이물질) 모두에서 동일 경향. → 데이터의 “출처(의미 수준 다양성)”가 “양”보다 결정적이라는 핵심 증거.

## 소규모 산업 결함 검출을 위한 생성형 AI 증강 효과

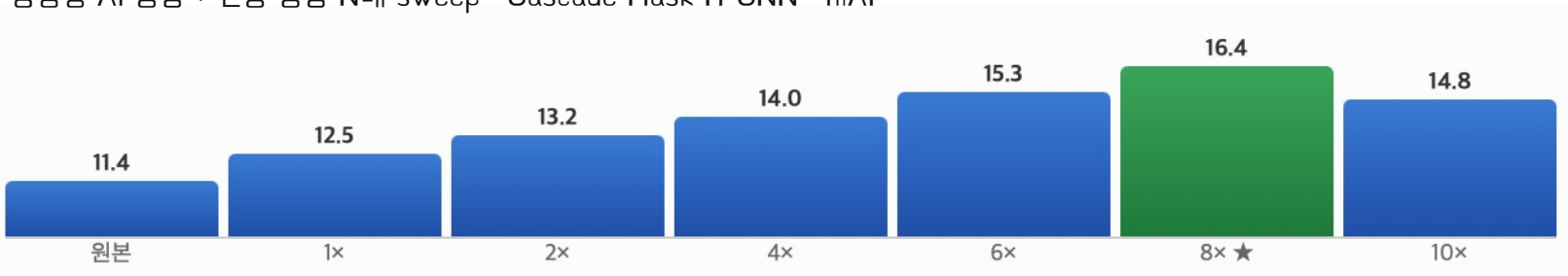
### 4. 실험 1 상세 - 모델·클래스별 결과 (표 1)

모델	클래스	원본	전통 증강	생성형 AI
Mask R-CNN	주조 개재물	12.4	10.28	14.1
	콘솔 오염	4.8	3.71	6.5
	목재 이물질	13.7	11.63	16.0
Cascade Mask R-CNN	주조 개재물	12.5	11.44	12.9
	콘솔 오염	7.9	6.88	8.7
	목재 이물질	13.3	12.76	13.8

6개 조건 전부에서 전통 증강 < 원본 < 생성형 AI 증강 패턴이 일관되게 재현. 우연이 아닌 구조적 차이임을 시사.

### 5. 실험 2 - 결합 시 시너지, 8배 지점이 최적

생성형 AI 증강 + 전통 증강 N배 sweep · Cascade Mask R-CNN · mAP



분량을 늘릴수록 성능 향상 → 8배 지점에서 최대 +5 mAP 향상 (원본 11.4 → 16.4). 그 이상 늘리면 모델이 원본 분포로 끌려가 과적합. 두 증강은 경쟁이 아닌 보완 관계: 생성형 AI = 의미 수준 다양성, 전통 = 픽셀 수준 미세 변형.

## 소규모 산업 결합 검출을 위한 생성형 AI 증강 효과

### 6. 실험 3 - 6개 모델 계열에서 일관된 효과

<b>Mask R-CNN</b> 2-stage CNN 결합 ✓	<b>Cascade Mask R-CNN</b> 2-stage CNN +5 mAP ✓	<b>SOLOv2</b> 1-stage CNN 결합 ✓
<b>RTMDet-Ins</b> 1-stage CNN 결합 ✓	<b>MaskDINO</b> Transformer 결합 ✓	<b>Mask2Former</b> Transformer 결합 ✓

2-stage CNN · 1-stage CNN · Transformer 세 계열의 대표 모델 모두에서 결합 조건 > 원본 조건. 트랜스포머 계열은 절대 성능은 낮지만 (데이터 의존성), 방향성은 동일 유지. → 제안 효과가 특정 구조에 국한되지 않음을 검증.

### 7. 실용적 함의 & 메시지

1. 실무 가이드: 더욱 많은 다중 시드로 실험을 해야 하지만 현재 실험 결과로서는 데이터가 부족할 때 전통 증강을 마구 늘리는 관행은 오히려 성능을 깎을 수 있다. 먼저 생성형 AI로 의미 수준 다양성을 확보한 후, 그 위에 전통 증강을 ~8x 결합하는 것이 본 실험의 최적 지점이다. 하지만 이는 절대적이 아닌 단순 실험결과 이다. 데이터 도메인마다 다를 수 있으며 절대적이 아니므로 추후 연구가 필요하다.

2. 학술적 메시지: 소규모 데이터에서는 "얼마나"보다 "어디서 온"이 중요하다 원본의 픽셀 변형을 넘어 의미 공간을 확장하는 증강이 핵심.

## 통합 전처리 파이프라인을 활용한 멀티모달 가스 센서 데이터셋에서의 모델 성능 비교

An Integrated Preprocessing Pipeline for Model Performance Comparison on a Multimodal Gas Sensor Dataset · MultimodalGasData 벤치마크 재현성 확보 + 8개 모델 공정 비교

### 1. 문제 인식-동일 데이터셋, 다른 전처리

MultimodalGasData를 활용한 기존 연구들은 전처리 절차가 모두 달라 공정한 성능 비교가 불가능하다 (Utomo et al., 2025).

a. 가스 분사 시점 정의 불명확 - 연구마다 anchor가 달라 같은 샘플도 다른 구간으로 잘림

b. 스케일링 누락 - 센서별 반응 범위 차이가 학습 신호를 오염, 모델이 "스케일"을 학습

c. 이상 구간 미제거 - 비정상 응답이 모델 평가를 왜곡, 재현성 불가

→ 전처리를 표준화하기 전까지, 어떤 모델이 진짜 우월한지 말할 수 없다

### 2. 핵심 명제 - 4-Stage 통합 전처리 파이프라인

가스 분사 시점 자동 정렬 → 이상 구간 제거 → 정규화 → 재현 가능한 교차검증 분할까지를 하나의 표준 파이프라인으로 통합.

S1 ▶ Anchor Detection :  $|\Delta S(t)| > \tau \rightarrow$  분사 시작/종료 자동 정렬

S2 ▶ Outlier Removal : 비정상 응답 구간 제거 (파형 기반)

S3 ▶ Min-Max Normalize :  $S' = (S - S_{\min}) / (S_{\max} - S_{\min})$

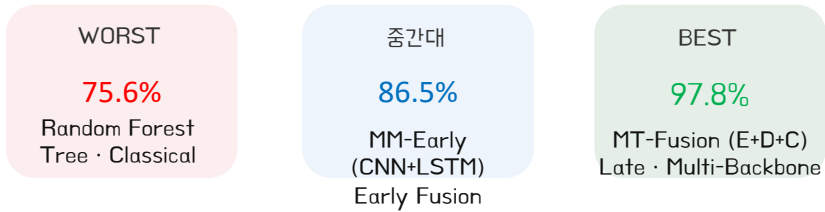
S4 ▶ Split 80/20 × 20 reps+ LSTM용 ±5 시점 윈도우

Output :  $\mathcal{D}_{\text{clean}} = \{ (X_{\text{train}}^{(k)}, X_{\text{test}}^{(k)}) \}_{k=1..20}$  · 모든 모델이 동일 split·동일 스케일 위에서 평가

## 통합 전처리 파이프라인을 활용한 멀티모달 가스 센서 데이터셋에서의 모델 성능 비교

An Integrated Preprocessing Pipeline for Model Performance Comparison on a Multimodal Gas Sensor Dataset · MultimodalGasData 벤치마크 재현성 확보 + 8개 모델 공정 비교

### 3. 동일 파이프라인 위 8-모델 Accuracy 격차



최상위와 최하위 사이 Accuracy 22.2%p 격차 같은 데이터에서 모델 선택이 결정적임을 정량 입증. MT-Fusion 두 모델만 90% 돌파.

### 4. 8개 모델 Accuracy 전체 비교 (20-rep CV 평균)

순위 · 모델	Accuracy	$\Delta$ vs Best
MT-Fusion (E+D+C)	0.978	
MT-Fusion (R+I+M)	0.922	-5.6 %p
CNN+LDA	0.900	-7.8 %p
4MM-Early (CNN+LSTM)	0.865	-11.3 %p
5LSTM	0.848	-13.0 %p
6CNN	0.847	-13.1 %p
7IDFC	0.810	-16.8 %p
8Random Forest	0.756	-22.2 %p

## 통합 전처리 파이프라인을 활용한 멀티모달 가스 센서 데이터셋에서의 모델 성능 비교

An Integrated Preprocessing Pipeline for Model Performance Comparison on a Multimodal Gas Sensor Dataset · MultimodalGasData 벤치마크 재현성 확보 + 8개 모델 공정 비교

5. 본질성 검증 : 융합 전략이 모델 깊이보다 중요

전략 (Strategy)	대표 모델	Acc 범위
Late · Multi-Backbone	MT-Fusion (E+D+C / R+I+M)	0.922 – 0.978
Early Fusion	MM-Early (CNN+LSTM), CNN+LDA	0.865 – 0.900
Single Modality	CNN, LSTM	0.847 – 0.848
Tree / Classical	IDFC, Random Forest	0.756 – 0.810

표준 전처리 위에서 비교했을 때, MultiTask Fusion만이 90% 정확도 벽을 넘는다. 모델 비교의 신뢰성은 데이터 파이프라인의 표준화에서 시작된다.

## 통합 전처리 파이프라인을 활용한 멀티모달 가스 센서 데이터셋에서의 모델 성능 비교

An Integrated Preprocessing Pipeline for Model Performance Comparison on a Multimodal Gas Sensor Dataset · MultimodalGasData 벤치마크 재현성 확보 + 8개 모델 공정 비교

가장 좋았던 모델 MT-Fusion (E+D+C) · Accuracy

성능이 좋았던 이유 | 백본 다양성이 만든 상호보완

EfficientNet(스케일링 효율), DenseNet(특징 재사용), ConvNeXt(현대화된 conv) 세 백본의 inductive bias가 서로 달라 동일 입력에서도 보완적 표현을 학습. 단일 CNN(0.847)이 포착 못한 미세 패턴까지 커버.

2Late Fusion으로 모달 정보 손실 최소화

센서(E-nose)와 열화상 이미지를 각 백본에서 충분히 학습한 뒤 후단에서 결합. Early Fusion(0.865)이 초기 단계에서 모달 특성을 섞으며 잃는 정보를, Late 구조가 끝까지 보존.

3멀티태스크 학습으로 미세 반응 차이 정밀 포착

"가스 종류 식별"과 "존재 여부"를 동시 학습해 태스크 간 지식 공유. NoGas vs Mixture vs Smoke vs Perfume처럼 반응 패턴이 유사한 클래스 사이의 경계가 뚜렷해짐 (AUROC 0.999).

결론. Late Multi-Backbone 구조가 우월했던 이유는 모델 깊이가 아니라 표현 다양성이다. 같은 데이터·같은 전처리 위에서, 서로 다른 inductive bias를 가진 백본들을 후단에서 합치는 설계가 가스별 반응 패턴의 미세 차이를 가장 잘 포착했다 이것이 22.2%p의 격차를 만든 본질이다.

## 퍼스널 컬러 진단 시스템

2024.06 - 2025.03

사용자의 퍼스널 컬러를 진단하고, 진단 결과와 정보에 맞는 영상 콘텐츠를 1인 부스 환경에서 체험할 수 있도록 한 **인터랙티브 웹**. (화성시 AI 초청)

<https://github.com/Woochang4862/personal-color-app>

*Python · PyTorch · AWS · Docker · react, java script, OpenAI, git . . etc*

---

## Car\_Nerf : AI 중고차 딜러

2025.12 – Now (경진대회 참가 중)

판매자가 차량 영상 1분만 업로드하면 **FastGS(CVPR 2026)** 기반 3D 모델링부터 **YOLOv8**외관 파손 진단·**LightGBM** 가격 예측 AI · 허위 매물 판별 · **LLM 자연어 검색**까지 전 과정을 자동화 해주는 중고차 거래 통합 웹/모바일 어플리케이션 시스템.

<https://github.com/Wlsgdh/CarNeRF>

*Python · TypeScript · JavaScript · Bash · FastAPI · SQLAlchemy · FastGS · COLMAP · SAM . . etc*

<http://Lifeai.suwon.ac.kr:5199>

---

## ETF with AI

2026.02 – Now (경진대회 참가 중)

AI 기반 ETF/주식 종목 랭킹 예측 + 자동 데이터 수집 파이프라인 + 웹 대시보드를 통합한 **ETF 트레이딩 시스템**

<https://github.com/Wlsgdh/etf-trading-projects>

*LightGBM LambdaRank (Learning-to-Rank) · FastAPI · Next.js · Playwright · MySQL · Docker Compose . . etc*

<http://ahnbi2.suwon.ac.kr/trading>

<http://ahnbi2.suwon.ac.kr>

---

## JUMP AI 2025: 제3회 신약개발 경진대회

2025.07 - 2025.08

Dacon 신약개발 경진대회 ; SMILES 입력으로 MAP3K5 표적 IC50 활성을 예측하는 신약 후보물질 스크리닝 모델

<https://github.com/Wlsgdh/Jump-AI-2025>

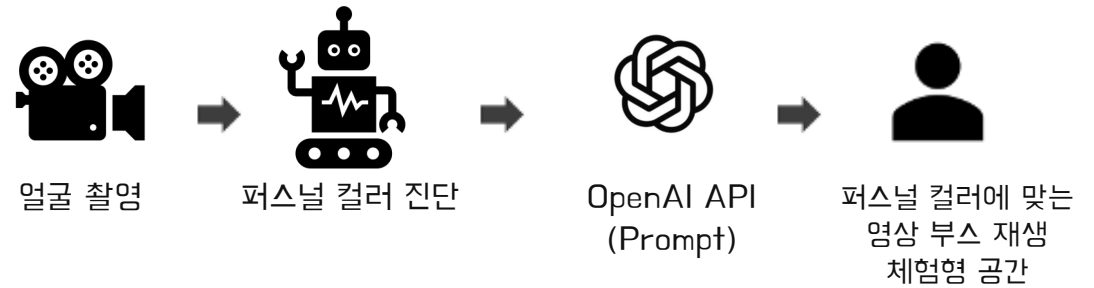
*Python · RDKit · scikit-learn · LightGBM · XGBoost · RandomForest · Extra Trees · Optuna . . etc*

---

## 퍼스널 컬러 진단 시스템

### 프로젝트 선정 이유

Computer Vision을 처음 공부하기 시작 했을 때, 했던 프로젝트로 단순 호기심에서 시작했습니다. 패턴인식에 최적화된 CNN이 과연 패턴인식 보다는 퍼스널 컬러의 단순 색 조합에서도 뛰어난 성능을 보이는지 궁금하여 시작하고, 성능이 생각보다 잘 나와 다른 비교군 모델을 제외 한 채 화성시 AI 포럼에서 발표하게 된 프로젝트입니다.



## 퍼스널 컬러 진단 시스템

### Phase 1 데이터 수집/정제

- 퍼스널 컬러가 알려져 있는 연애편들에 대해 구글 이미지 데이터 수집 약 5만장
- 화이트 스케일링을 진행하여 조명 영향을 많이 받은 사진들을 정제

### Phase 2 Segmentation

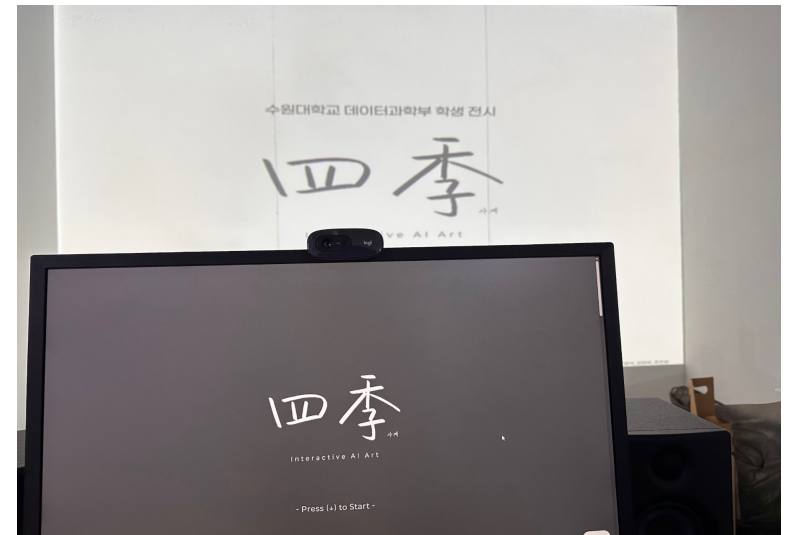
- OpenCV로 눈,코,입을 제거 후 피부 톤을 모델 학습에 사용

### Phase 3 LLM

퍼스널 컬러에 맞는 분위기와 생성형AI를 통해서 프롬프트를 전달 받고 이에 맞는 영상을 선택

### Phase 4 부스 연결

LLM에게 받은 결과를 실시간으로 전달 받자마자 영상 및 사운드 재생

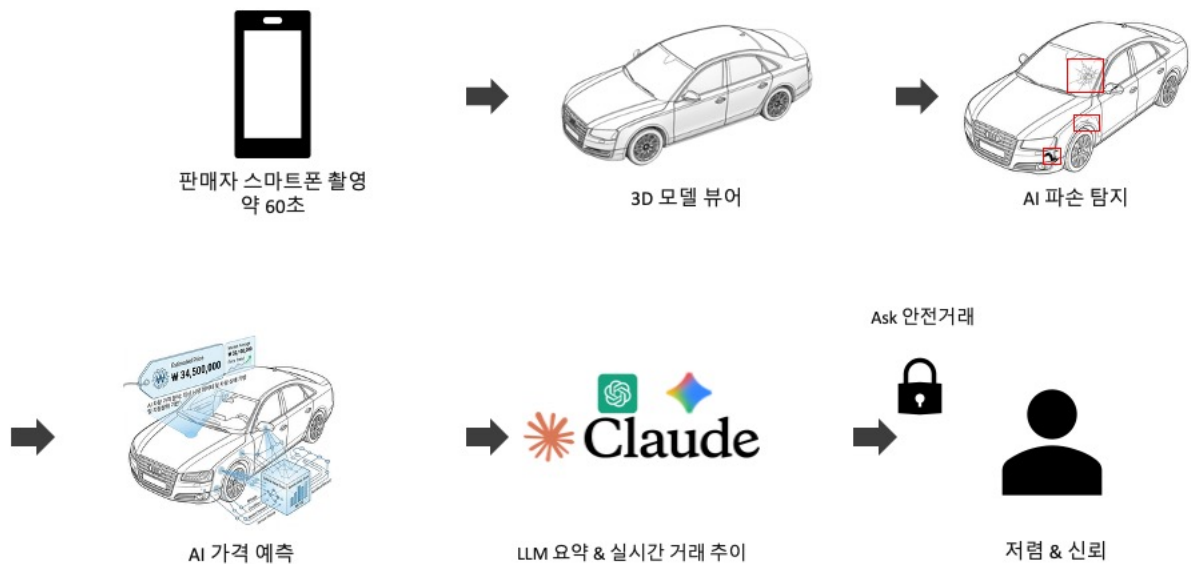


## Car\_Nerf : AI 중고차 딜러

## 프로젝트 선정 이유

Nerf라는 논문을 알게 되었고, 부동산에서 실질적으로 활용되는 예시를 보고, 호기심과 흥미를 가지고 3d gaussian splatting 논문까지 읽고 이를 직접 구현 하고 싶다고 생각했습니다. 이후 CVPR 2026에서 fastGS가 나오고 이를 직접 중고차 시장에서 활용할 수 있다는 생각에 AI 딜러 Car\_Nerf를 시작하게 되었습니다.

## 신뢰있는 AI 중고차 딜러



## Car\_Nerf : AI 중고차 딜러

### 신뢰있는 AI 중고차 딜러

#### Phase 1 기획 & 백엔드 기반 구축

- 중고차 3D 모델링 + AI 결함 탐지 플랫폼 컨셉 정립
- FastAPI + SQLite + Cookie JWT 인증 구조 설계
- DB 모델링
- 16개 API 도메인 구축

#### Phase 2 3D 파이프라인 완성

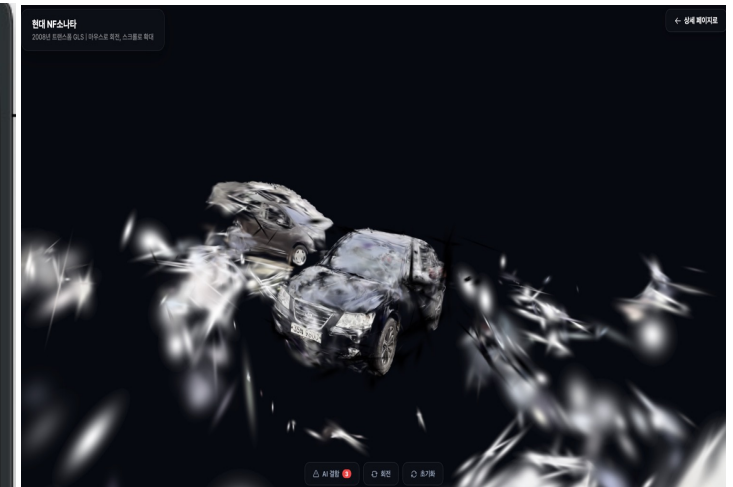
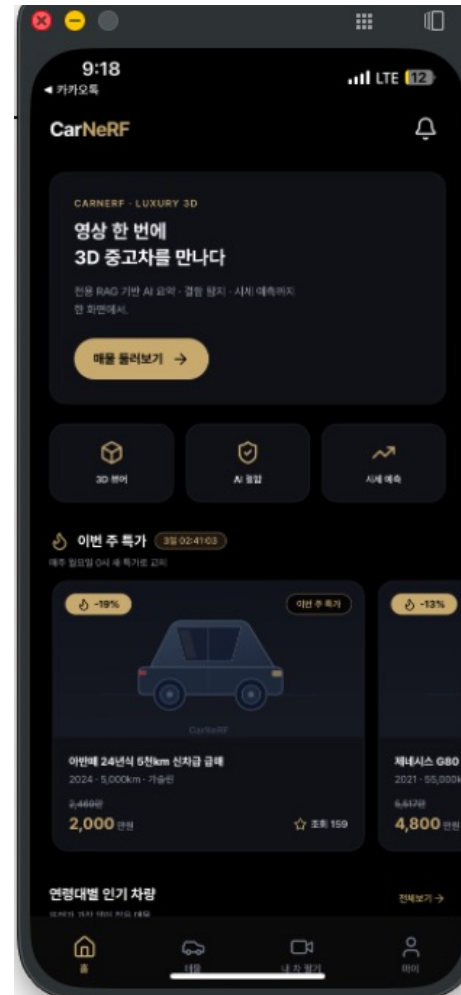
- extract\_frames → COLMAP → rembg + depth → Gaussian Splatting → export → web 폴 자동화
- 초기 vanilla 3DGS (60K iter) 기반 학습 → PSNR 31.88 달성
- pycolmap segfault 이슈를 OPENBLAS\_NUM\_THREADS=1로 해결, depth\_params 자동 비활성화 등 다수의 옛지케이스 해결

#### Phase 3 AI 모듈 통합

- YOLOv8 단일 클래스 차량 결함(damage) 탐지 모델 학습 (phase1\_r3, YOLO26l-seg)
- LightGBM 기반 가격 예측 모델 v2 + predict.py 개선, sell 페이지 AI 시세 연동
- 전용 RAG 기반 차량 요약(ChatGPT) 모듈 마케팅 핵심 차별화 포인트

#### Phase 4 3D 엔진 고도화 (FastGS 마이그레이션) + 모바일 확장

- vanilla 3DGS → FastGS (CVPR 2026) 전면 교체
- 30K iter ≈ 2~3분 (A100, 기존 60K 대비 15배 가속)
- fastgs 전용 conda 환경 (Py3.7 + CUDA 11.8) 분리, PLY 포맷 호환성 유지
- 2DGS / MCMC 변형 옵션 엔진으로 보존
- React Native(expo) 기반 초기 베타모델 구현



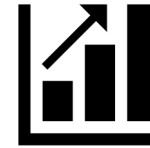
ETF with AI

프로젝트 선정 이유

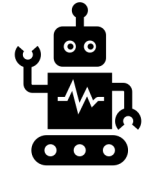
Lightgbm LambdaRank 라는 모델을 알게 된 이후로, 이를 직접 적용해보고 싶은 계기가 되었고, 랭킹 시스템에서 영화 추천 알고리즘은 이미 이전에 했던 경험이 있어 AI가 직접 ETF종목을 모을 수 있으면 좋겠다고 생각하여 시작했습니다.



사용자 ETF  
분석 요청



모델이 ETF  
100종목 추천



100일(거래63일)  
매수/매도  
자동화



감정 매수/매도  
기록



시장/펀더멘탈 데이터  
제공 및 LLM 기반  
종목 분석

## ETF with AI

### Phase 1: 기초 인프라 & ML 실험

#### 프로젝트 뼈대 구축 + ML 모델 비교 실험

- Docker + Nginx + FastAPI 기반 ML 서비스 초기 구현
- ML 모델 비교 실험 (Ridge, Lasso, RandomForest, TabPFN 등) → 앙상블 시도
- GitHub Actions CI/CD 파이프라인 + 팩트시트 PDF 자동 생성

### Phase 2: 데이터 수집 파이프라인

#### TradingView 실시간 주가 자동 수집

- Playwright 기반 TradingView 스크래퍼 개발 (OHLCV + 기술지표)
- SSH 터널을 통한 원격 MySQL 연동
- Auto-Monitoring 대시보드 구축

### Phase 3: AhnLab 모델 & 피쳐 엔지니어링

#### 프로덕션 ML 모델 (LightGBM LambdaRank) 구축

- 85개 피쳐(기술지표 + 거시경제 + Z-score) 기반 LambdaRank 모델 개발
- 전처리 DB 분리, Poetry 패키지 관리 전환

### Phase 4: 자동매매 시스템

#### KIS API 연동 실패매 + 모니터링

- trading-monitor Next.js 대시보드 신규 개발
- 자동매매 파이프라인 완성 (스크래핑 → 피쳐처리 → ML 예측 → KIS 주문)
- KIS API 연동 (잔고 / 주문 / 체결 / 토큰 자동 관리)
- 모니터링 기능 다수 구축 (매매 달력, 수익률, DB Viewer, 환율 계산기 등)
- cron 자동화 (매일 예측 + 월간 재학습)

### Phase 5: 멀티시 플랫폼 & 커뮤니티

#### 멀티시 융합 + 커뮤니티 + 스크리닝 기능 확장

- 멀티시 융합 플랫폼 (Technical / Fundamental / Market 3x3 그리드)
- 스크리닝 페이지 + 관심종목 기능, 종목 1000개 확장



종목 랭킹 (최신 예측) 966종목 전체 보기

#	종목	점수	현재가	방향
1	RCKT	0.9301	\$3.60	BUY
2	LCID	0.8468	\$8.77	BUY
3	MARA	0.7975	\$9.50	BUY
4	BBAI	0.7970	\$3.49	BUY
5	BTDR	0.7728	\$10.36	BUY
6	ARRY	0.7542	\$7.30	BUY
7	LUNR	0.7256	\$21.97	BUY
8	CLSK	0.7228	\$9.88	BUY
9	SNAP	0.7070	\$4.95	BUY
10	LIT	0.6577	\$2.26	BUY
11	SOUN	0.6321	\$6.79	BUY
12	PATH	0.6160	\$10.39	BUY
13	ANGI	0.5946	\$7.20	BUY
14	AI	0.5892	\$8.78	BUY

## JUMP AI 2025: 제3회 신약개발 경진대회

### JUMP AI 2025 경진대회

머신러닝 프로젝트의 경험을 쌓고 하이퍼파라미터 튜닝 경험을 쌓기 위하여 경진대회에 참가하였으며, 1134명 중 개인 참여로 20위 최종 마감을 하였습니다.

### Jump AI(.py) 2025 : 제 3회 AI 신약개발 경진대회

알고리즘 | 분자 구조 | 정형 | 회귀 | 바이오

Ⓜ 상금 1,500만 원

🕒 2025.07.07 ~ 2025.08.25 09:59 [+ Google Calendar](#)

👤 1,134명 📅 마감



[사블록북했어요!](#)

[참여중](#)

[대회안내](#) [데이터](#) [코드 공유](#) [토크](#) [리더보드](#) [제출](#)

- 개요**
- 규칙
- 일정
- 상금
- 동의사항

**[배경]**  
AI 신약개발 생태계 활성화와 AI 신약 개발 젊은 연구원들의 의욕 고취 및 인재 유입을 위해 'Jump AI 2025 : 제 3회 AI 신약개발 경진대회'를 개최합니다.

**[주제]**  
MAP3K5 IC50 활성값 예측 모델 개발

**[설명]**  
PubChem, ChEMBL, CAS 등에서 수집한 실험 기반 화합물 정보를 기반으로 IC50값 예측모델 개발  
127종 화합물의 구조 정보를 입력값으로 사용하여 해당 화합물들의 ASK1에 대한 IC50 값을 예측하여 제출

**[주최 / 주관 / 운영]**

- 주최/주관 : 한국제약바이오협회
- 후원 : 보건복지부, 유한양행, CAS
- 운영 : 데이콘

**[참가 대상]**  
전국민 누구나

## JUMP AI 2025: 제3회 신약개발 경진대회

### PHASE 1 · Target 변환 & 데이터 정리

- IC50 → pIC50 변환 ( $pIC50 = 9 - \log_{10}(IC50 [nM])$ )으로 학습 공간 정규화
- IC50 ∈ [0.1, 1e5] nM 범위로 필터링

### PHASE 2 · Feature Engineering

- RDKit Descriptors 약 40종 추출
- Morgan Fingerprint 1024-bit (ECFP4, radius=2) 추출 → PCA로 100차원 압축
- 최종 입력:  $X = [RDKit(40) \parallel PCA(Morgan FP, 100)] \in \mathbb{R}^{140}$

### PHASE 3 · 전처리 & HPO

- 3-Scaler 비교 (Standard / Robust / Quantile) → RobustScaler 채택
- Optuna 기반 HPO 수행 (30 trials × 5-fold CV, RMSE objective)
- 정규화 파라미터 탐색 (reg\_alpha, reg\_lambda, min\_child\_\*)

### PHASE 4 · 5개 Base Learner 학습

- LightGBM (leaf-wise) / XGBoost (level-wise)
- Random Forest / Extra Trees
- MLP (256-128-64)
- 20% held-out validation 기반 early stopping, 동일 피쳐 + RobustScaler 적용

### PHASE 5 · Two-Track Ensemble Blending

- Track A · SLSQP 가중 평균 min\_w RMSE 가중평균 + rank-average 결합
- Track B · Quantile Matching 모든 모델 예측을 RF 분위수에 매핑 후 평균
- Meta-Blend: Final = 0.6 · A + 0.4 · B

### PHASE 6 · 후처리 & 제출

- pIC50 예측을 [y\_train\_min, y\_train\_max]로 1차 클리핑
- IC50 = 10^(9 - pIC50)로 nM 단위 역변환
- 최종 IC50를 [0.1, 1e5] nM로 2차 클리핑

1	koolss		0.75069	51	9달 전
2	도라에몽		0.73928	121	9달 전
3	파이썬수업너무어렵다		0.70252	54	9달 전
4	노아		0.69479	56	9달 전
5	Ghent_AI		0.69407	130	9달 전
6	jt kim		0.68854	105	9달 전
7	Agnes		0.66764	56	9달 전
8	진		0.66352	38	8달 전
9	happyyyy		0.65809	38	9달 전
10	IIS		0.65533	19	9달 전
11	code7monkey		0.6543	49	9달 전
12	성대_의약화학		0.65347	65	8달 전
13	TADA		0.65105	29	9달 전
14	ASKME		0.64527	47	9달 전
15	Leeee		0.64461	40	8달 전
16	asdf98		0.62542	22	10달 전
17	은비비빅		0.62432	96	9달 전
18	수면도사		0.62153	63	9달 전
19	비비고곰탕		0.62142	19	9달 전
20	브로투푸수		0.6167	14	9달 전
21	3시에퇴근하는마법의소라고동		0.61458	63	9달 전
22	B105-1		0.61382	46	9달 전

## EXPERIENCES

---

### 대한민국 해병대

2022.01 - 2023.07

---

### AIMMO 주식회사 · 자율주행 부서 데이터 수집/정제

2024.09 - 2024.12

Vision Fail 데이터 정제, fine-tuning, 신규 데이터 수집 업무 수행.

---

### 수원대학교 학부 연구생

2024.09 - 현재

Vision 분야 연구 참여. React · Node.js · MongoDB 기반 응용 프로젝트 진행.

---

### 성균관대학교 인턴 · Hadd Science 인턴

2025.09 - 2025.12

Hadd Science 사이트 관리, 데이터 관리, 기사 작성, Auto Cad 작업

---

### 수원대학교 DSML Vibe Coding 학생 조교

2025.09 - 현재

학부생 대상 코딩 수업 운영 보조 및 실습 멘토링 진행.

---

### GUIP 바이오헬스 플랫폼 활동 수료

2025.11 ~ 2025.12

디지털 헬스케어 초급 수료

---

### 구글 코세라 데이터 준전문가

2025.02

Google Data Analytics Professional Certificate 취득.

---

# EXPERIENCES


문서확인번호 : 1778-7569-9885-4335

■ 병역법 시행규칙 (별지 제5호서식)      상봉과 (www.gov.kr) 에서도 신청할 수 있습니다. 1/1

### 병적증명서


발행번호	182807	유효기간	무기한
종 도	[ ] 공적지 등 신고용 [ V ] (경력 확인)		
연적사항	성명    주진호      생년월일    2002. 07. 09.		
군(대체) 병역 여부	[ V ] 복무를 마친 사람      [ ] 복무를 마쳐지 않은 사람		
병역사항	<ul style="list-style-type: none"> <li>○ 군별                    해군(해병대)</li> <li>○ 계급                    병장</li> <li>○ 군번                    227200099</li> <li>○ 직종                    제1차</li> <li>○ 입영(입관)/소집일자    2022. 1. 24.</li> <li>○ 전역(소집해제)일자    2023. 7. 23.</li> <li>○ 전역(소집해제)구분(사유)    전가</li> </ul> <p>- 이 외 여 액 -</p>		

'병역법 시행규칙' 제8조에 따라 위와 같이 병역을 증명합니다.      2026년 05월 14일

경인지방병      

● 본 증명서는 인터넷으로 발급받으며, 증명서(www.gov.kr)에 인터넷공인인증서(개인)의 사용을 통해 위·변조 여부를 확인할 수 있습니다. \*복합문자까지 999까지 포함 문서 확인의 이유로도 전자확인(정부24 및 본인 소지내용 문서확인 프로그램)을 활용 수 있습니다.

경기도 성남시 분당구 판교로 228번길17  
한국과학기술원(GIST) 2호관 505호 5층  
TEL:031-706-3533 FAX:070-8255-3389




## 경력 증명서

성 명	주진호	생년월일	2002.07.09
사업장주소	경기도 성남시 분당구 판교로228번길 17 한국세분벤처밸리2 이면책용 9층		
소 속	AIMMO 자율주행팀		
업 무 내 용	자료수집 및 영상분석		
재 직 기 간	2023.08.03 ~ 2024.02.29		

상기 내용은 사실과 다름 없음을 증명합니다.

2024년 02월 29일

췌 예이모      

대표이사    오 승 탕

제 2025-25호

## 수료증

데이터과학부  
21016094  
주진호

위 학생은 수원대학교 GUIP 바이오헬스 플랫폼 비교과프로그램의 모든 과정을 성실히 이수하였으므로 본 수료증을 수여합니다.

1. 디지털 헬스케어의 이해와 사례
2. 디지털 헬스케어 세포 매양 심층
4. 디지털게어를 위한 유전체 데이터 처리
5. 디지털게어를 위한 유전체 데이터 AI모델링 및 활용

교육금수: 초급  
교육기간: 2025년 11월 19일~12월 17일  
\*본 프로그램은 경기도 대학혁신공용협의체 지원을 받아 수행되었습니다

2025년 12월 30일  
수원대학교        
GUIP사업단 이 유 현      



## 인턴 수료증


성 명	주진호	생년월일	2002년 07월 09일
소 속	데이터관리부	직 위	인턴

상기인은 2025년 08월 25일부터 2025년 11월 26일까지 당사 하드웨어에 인턴으로 근무하여 수료하였음을 증명합니다.



종 도 : 외부기관 제출용

2025년 12월 01일


주 소 : 경기도 화성시 봉담읍 와우안길 17, 수원대학교 고운첨단과학기술원 514호

회 사 명 : 하드사이언스      

대 표 이 사 : 허재정

Feb 27, 2025

진호주      

has successfully completed the online Professional Certificate

## Google Advanced Data Analytics

Those who earn the Google Advanced Data Analytics Certificate have completed seven courses that include hands-on, practical, case-based assessments and are designed to prepare them for advanced roles in data analytics and entry-level roles in data science. They are competent in exploring large datasets, applying data analysis techniques, and building models to extract insights. They are also competent in machine learning, predictive modeling, and statistics.

Foundations of Data Science  
Get Started with Python  
Go Beyond the Numbers: Tackling Data with Analytics  
The Power of Statistics  
Regression Analysis: Singlely  
Complex Data Relationships  
The Nuts and Bolts of Machine Learning  
Google Advanced Data Analytics Capstone

Welly this certificate at: <https://www.coursera.org/verify/certificate/21016094>

*저의 장점을 소개합니다 !*

## 호기심에서 시작되는 연구

배움에 있어 얕은 이해에 머무르지 않고, 떠오른 궁금증을 끝까지 파고드는 것을 좋아합니다. 직접 가설을 세우고 작은 실험부터 설계해 검증해 나가는 과정에서 자발적인 연구 경험을 쌓아왔으며, 그 과정에서 마주한 수많은 실패 또한 더 깊이 있는 탐구로 이어지는 값진 경험이 되었습니다. 이러한 경험들은 대학원 연구 환경 속에서도 스스로 질문을 던지고, 끈기 있게 답을 찾아 나가는 연구자로 성장하는 든든한 밑거름이 될 것이라 확신합니다.

---

## 문제 해결 능력

평소 창업 활동 도전에 관심이 많아, 주변 사람들이 겪는 불편함을 그냥 지나치지 않고 어떻게 해결할 수 있을지 고민하는 습관을 가지고 있습니다. 문제 앞에서 피하기보다 정면으로 마주하는 편이며, 기술적 설계를 진행할 때에도 끊임없이 "왜?"라는 질문을 던지며 현상이 아닌 본질을 파악하려 노력합니다. 이러한 사고방식은 복잡한 SI 모델을 적용하거나 새로운 논문을 읽을 때에도 자연스럽게 이어져, 어떤 개념도 어물쩡 넘기지 않고 깊이 있게 이해하려는 자세로 연결되었습니다. 앞으로의 연구에서도 이러한 태도를 바탕으로 본질에 닿는 해결책을 끈기 있게 찾아 나가고자 합니다.

---

## 팀과 함께하는 성장

개인적인 호기심에서 출발한 연구를 제외한 대부분의 프로젝트에서 팀장 역할을 맡아왔습니다. 팀 프로젝트에서 가장 중요하게 생각하는 것은 각자의 강점을 살린 역할 분담과, 그 위에서 만들어지는 유기적인 협업입니다. 개인의 역량을 단순히 합치는 것이 아니라, 구성원 간의 원활한 소통을 통해 더 큰 시너지를 만들어내는 과정이 곧 팀 프로젝트의 본질이라 믿습니다. 이러한 경험을 통해 어떤 조직에서도 협력과 리더십을 발휘할 수 있다는 자신감을 얻었으며, 앞으로의 연구 환경에서도 동료들과 함께 배우고 성장하며 더 깊이 있는 결과를 만들어가는 연구자가 되고자 합니다.

## VISION

*끊임없는 연구와 학습을 통해 신뢰받는 연구자로 성장하겠습니다.*

## 인턴의 자세 및 목표 다짐

---

### 의사소통을 통한 연구 효율의 극대화

연구는 결국 사람과 사람 사이의 협업으로 완성된다고 믿습니다. 정기적이고 자발적인 보고 미팅을 통해 진행 상황과 어려움을 투명하게 공유하고, 교수님 및 동료 연구자들과 과제의 방향성을 함께 점검해 나가겠습니다. 또한 코드 공유와 피드백 과정을 체계화하여 작은 시행착오도 팀 전체의 자산이 될 수 있도록 만들고, 이를 통해 연구실 전반의 효율성과 결과물의 품질을 한 단계 끌어올리는 데 기여하겠습니다.

### 맡은 연구 과제에 대한 책임감과 리더십

제게 주어진 연구 과제는 그 누구보다 성실하게, 그리고 끝까지 책임지고 완수하겠습니다. 단순히 주어진 일을 처리하는 수준을 넘어, 과제의 배경부터 목적, 기여점까지 완벽히 이해하고 그 위에서 새로운 가능성을 제안할 수 있는 연구자가 되겠습니다. 나아가 팀 프로젝트 안에서는 동료들의 강점을 살리는 리더로서, 모두가 함께 목표에 도달할 수 있도록 적극적으로 이끌어 나가겠습니다.

### 질문을 두려워하지 않는 열린 학습 자세

모르는 것을 부끄럽게 여기기보다, 모르는 것을 인정하고 묻는 자세가 진짜 성장의 출발점이라고 생각합니다. 선배 연구자분들과 교수님께서 주시는 피드백을 한 마디도 흘려보내지 않고 노트에 기록하고 곱씹으며 제 것으로 만들어 가겠습니다. 항상 열린 자세로 배우되, 받은 가르침을 단순히 수용하는 데 그치지 않고 후속 질문과 직접적인 실험으로 이어가며 깊이를 더하겠습니다. 그렇게 쌓인 배움을 통해 연구실에 실질적으로 보탬이 되는 인력으로 성장하겠습니다.

### 연구실의 한 구성원이 되기 위한 진심

위의 모든 다짐은 결국 하나의 마음에서 출발합니다. 이 연구실의 일원이 되어, 함께 의미 있는 연구를 만들어가고 싶다는 진심입니다. 부족한 점이 있다면 누구보다 빠르게 채우고, 가진 강점이 있다면 연구실을 위해 아낌없이 쏟아붓겠습니다. 단순히 학위를 받기 위한 학생이 아니라, 교수님과 선배·동료들과 함께 성장하며 연구실에 긍정적인 흔적을 남기는 사람이 되고 싶습니다. 그 첫걸음을 이곳에서 시작할 수 있기를 진심으로 바랍니다.