상황이 바뀌어도, 함께 해결하는 AI

상황 변화에 따른 **재판단**과
행동 AI **파인튜닝**을 함께 검증합니다.

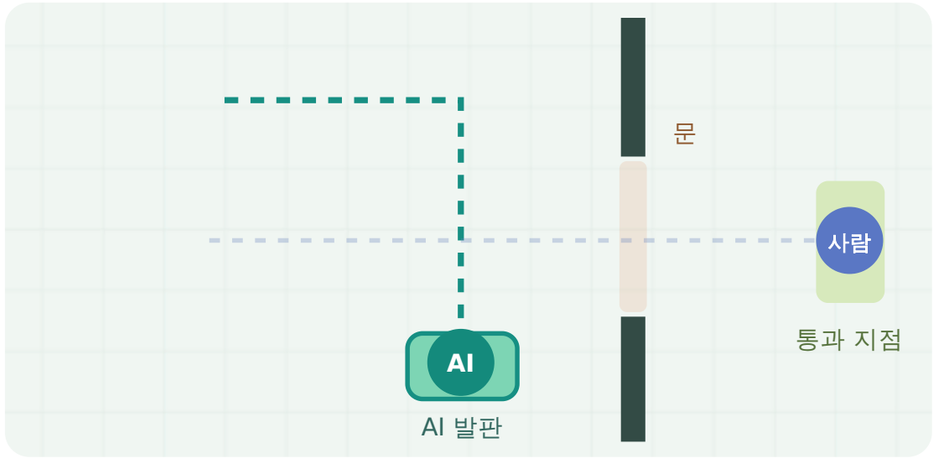
> **확정한 연구 방향**
>
> DPT-Agent의 실행·추론 분리 구조를 참고해 **상황 기반 재판단과 행동 AI 파인튜닝**을 적용하고, 협동 성공률·반응 시간·호출 비용을 비교한다.

조정: 언제 다시 판단할까 **학습: 무엇을 선택할까**

게임콘텐츠캡스톤디자인 · 1팀 더미
이영호 · 방경민 · 서준영 / 2026.10.02

연구 방향 확정 · 실제 모델 통합·학습·비교 실험은 진행 예정

사람과 AI가 함께 여는 문 개념 시연



발판 **유지 중** 문 **열림** 과제 **완료 확인**

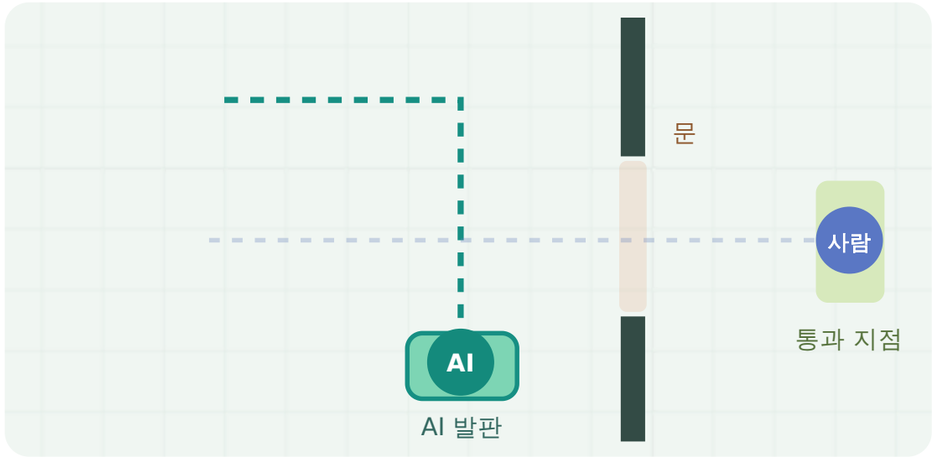
⑤ 완료 확인: 실제 사람 통과까지 확인해야 과제가 끝납니다.

설명을 위해 만든 자동 애니메이션입니다. 실제 게임·모델 실행 결과가 아닙니다.

발판에 도착하는 것과 함께 통과하는 것은 다릅니다.

● 사람과 AI가 함께 여는 문

개념 시연



발판 **유지 중** 문 **열림** 과제 **완료 확인**

⑤ 완료 확인: 실제 사람 통과까지 확인해야 과제가 끝납니다.

설명을 위해 만든 자동 애니메이션입니다. 실제 게임·모델 실행 결과가 아닙니다.

우리의 목표는 자연어 지시를 받아 **사람과 함께 과제를 끝내는 동료 AI**입니다.

1. **사람의 목표를 받는다**
   “내가 지나갈 때까지 발판을 밟고 있어 줘.”
2. **상황에 맞는 행동을 고른다**
   단순 이동이 아니라 발판을 계속 유지하는 행동이 필요합니다.
3. **실제 결과까지 확인한다**
   문이 열린 뒤 사람이 통과한 것을 확인해야 과제가 끝납니다.

게임을 실험 환경으로 쓰는 이유

상태·행동·성공 조건을 정하고, 같은 상황을 반복할 수 있습니다. 지시 변경이나 경로 차단도 통제해 넣을 수 있습니다.

판단하는 동안에도 게임과 사람은 움직입니다.

관측 → 판단 → 실행 사이의 간격

개념 시연

모델이 받은 상태

길이 열려 있음

AI ——— 발판

관측 시점

행동 선택 중

그동안 바뀐 실제 상태

길이 막혔음

AI — 장애물 — 발판

실행 시점

이전 상태의 응답

“발판으로 이동”

현재 상태로 재검사

④ 실행 전에 다시 검사합니다. 목표 관련 변화라면 재관측·재판단을 검토합니다.

개념 시간선입니다. 구간 길이는 실제 지연을 나타내지 않습니다.

언제 다시 묻고, 무엇을 다시 고를까요?

사람이 지시를 바꾸거나, 길이 막히거나, 행동이 실패하면 처음의 선택을 그대로 이어갈 수 없습니다.

너무 자주 판단하면

안정적인 상황에서도 모델을 호출해 시간과 토큰을 사용합니다.

너무 늦게 판단하면

변화에 대응하는 시점이 늦어지거나 유효하지 않은 선택이 남습니다.

이 절충 관계는 실험으로 확인할 대상입니다. 특정 방식의 우월성을 미리 가정하지 않습니다.

조정기는 호출 시점을,
학습은 행동 선택을 다룹니다.

축 1 · 조정

언제 다시 판단할까?

목표와 관련된 상태 변화가 생겼을 때 행동 AI를 선택적으로 다시 호출합니다.

변화 감지 → **재관측** → **같은 AI 재호출**

비교할 것: 성공률과 호출 비용의 관계

축 2 · 학습

무엇을 선택할까?

공개된 상태와 지시에서 이동·유지·대기·확인 질문을 잘 선택하도록 행동 AI를 파인튜닝합니다.

검토된 사례 → **LoRA 학습** → **선택 비교**

비교할 것: 새 퍼즐에서의 선택과 협동 성능

검증 질문 같은 기반 모델에서, **재판단 방식**과 **과제에 맞춘 학습**은 협동 성능·반응 시간·호출량에 각각 어떤 영향을 주는가?

조정기·기억·실행기는 일반 코드입니다. 파인튜닝은 기존 행동 AI의 가중치를 바꾸며 AI를 하나 더 추가하지 않습니다.

기준 연구를 정하고,
비교할 부분을 좁혔습니다.

ACL 2025

DPT-Agent

빠른 규칙 실행과 비동기 추론을 결합해 실시간 인간–AI 협동을 다룹니다.

가져오는 원리

고수준 판단과 지속적인 동작 실행을 분리합니다.

우리의 비교 초점

같은 실행기에서 재판단 호출 시점을 바꿔 성능과 비용을 측정합니다.

학회 원문 ↗

원 논문 2021

LoRA

기반 가중치를 고정하고 작은 변경분을 학습해 모델을 과제에 맞춥니다.

가져오는 방법

행동 AI 한 자리에 LoRA를 적용하는 방식을 우선 검토합니다.

우리의 비교 초점

동일한 기반 모델의 학습 전·후를 새 퍼즐에서 평가합니다.

논문 원문 ↗

이번 프로젝트의 기여 후보

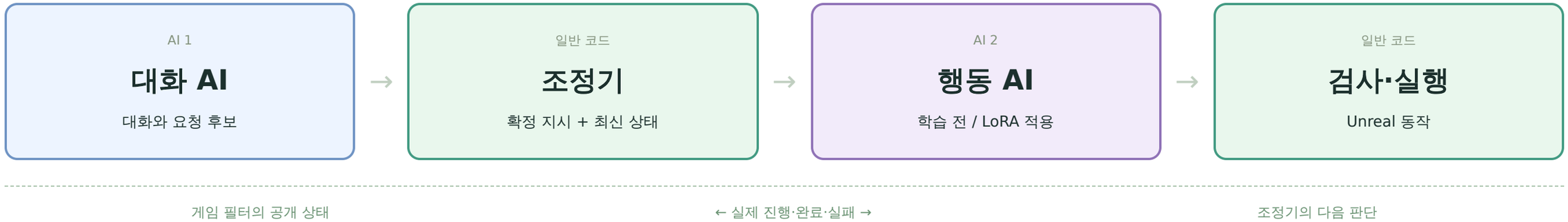
관측 정보와 상황 변화를 명세한 협동 환경에서, 호출 정책과 학습의 효과를 분리해 재현 가능한 비교 결과를 남기는 것입니다.

DPT-Agent도 환경 변화에 반응합니다. 역할 분리·상황 대응·LoRA 자체를 새로운 기법으로 주장하지 않습니다. 원 논문 전체 재현이나 우월성 비교는 별도 검증이 필요합니다.

AI는 두 개입니다.
조정기가 흐름을 이어 줍니다.

한 번의 요청이 실행 결과로 돌아오는 과정

설계 흐름



⑤ 실제 진행·완료·실패 결과를 조정기가 받아 다음 판단에 사용합니다.

지시를 확정한다

음성 인식 중간 결과나 침묵을 실행 동의로 처리하지 않습니다. 확정된 요청을 사용합니다.

실제 상태로 검사한다

오래된 요청·중복 실행·허용되지 않은 행동을 코드가 검사합니다.

실제 결과를 받는다

명령 접수와 과제 성공을 구분합니다. 실행 결과가 불명확하면 먼저 상태를 확인합니다.

기억은 관련 대화·확인된 결과를 보관하는 코드입니다. 실시간 상태를 파인튜닝 가중치로 대신하지 않습니다.

무엇을 미리 알려 주었는지 실험 조건으로 공개합니다.

01

공개 관측

관측 구역에 있는 객체와 현재 상태

- 객체 ID·역할·위치
- 발판 활성·문 상태
- 현재 제공되는 행동과 실행 결과

모든 주 비교 조건에서 동일

02

사전 설정 정보

개발자가 작성한 역할과 관계

- 어떤 상호작용을 지원하는가
- 발판과 문의 관계를 알려 주는가
- 후보 목록에 어떤 정보가 담기는가

제공 여부를 명시하고 고정

03

채점용 정답 상태

엔진이 내부 판정에 사용하는 정보

- 숨겨진 퍼즐 해법·관계
- 미래 결과와 평가 전용 정답
- 실제 통과·접촉·완료 기록

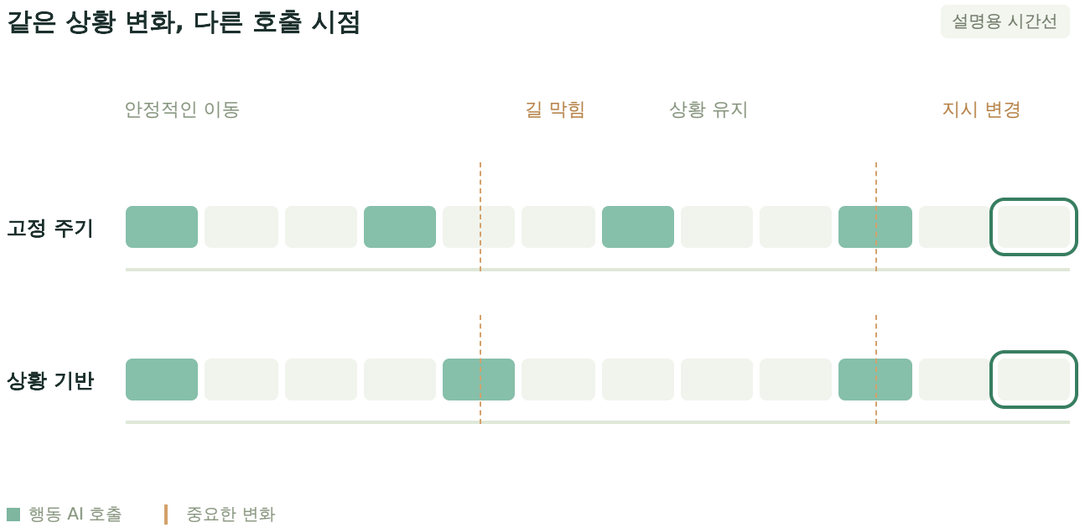
허용되지 않은 정보는 입력에서 제외

JSON으로 추출하는 것과 화면을 보고 알아내는 것은 다릅니다.

현재 방식은 설정된 객체 의미와 실행 중 상태를 구조화해 전달합니다. 임의의 게임 코드를 자동으로 이해하거나 영상에서 모두 인식한 결과로 설명하지 않습니다.

관계 정보의 양을 바꾸는 실험은 후속 비교입니다. 우선 재판단과 학습 효과를 볼 때는 입력 정보 범위를 고정합니다.

상황의 작은 변화마다 모델을 부르지는 않습니다.



호출을 검토할 사건

새 지시 확정
사용자가 목표나 대상을 바꿈

선행 조건 변화
목표 행동에 필요한 길·객체 상태가 바뀜

확인된 실행 실패
엔진이 실패를 보고해 다음 선택이 필요함

정상 이동 중의 작은 좌표 변화는 호출 조건에서 제외합니다. 반복 사건을 합치고 호출 상한을 둡니다.

두 방식의 공통 조건
실행 검사·취소 확인·다음 하위 행동 처리·관측 주기는 같게 두고, 호출 시점만 비교합니다.

모델 전체를 새로 만들지 않고, 행동 선택을 과제에 맞춥니다.

LoRA를 적용하는 위치

학습 원리

학습 중 고정

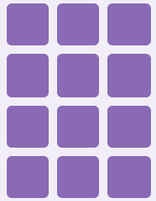
기반 모델

기존 지식과 언어 능력

+

학습할 부분

작은 변경분



과제에 맞는 선택

검토된 입력 → 올바른 행동·질문 사례

↑ 학습

같은 행동 AI 자리에 적용

기반 모델 + 학습한 변경분

④ 기반 모델과 학습한 변경분을 같은 행동 AI 자리에 적용합니다.

학습할 판단의 예

이동과 유지 구분

“여기로 와”와 “지나갈 때까지 밟고 있어”에 맞는 행동을 구분합니다.

문맥에 맞는 대상 선택

허용된 대화와 관측에서 “아까 말한 발판”이 무엇인지 판단합니다.

정보 부족 시 확인 질문

보이는 정보만으로 정할 수 없다면 필요한 질문을 선택합니다.

LoRA 원 논문 ↗ · PEFT 문서 ↗ — LoRA 자체가 모델을 작게 만들거나 1초 이하 응답을 보장하지 않습니다. QLoRA는 학습 메모리가 제약일 때 검토할 후보입니다.

게임 상태와 대화에서 선택까지 이어지는 사례를 만듭니다.

학습 사례 한 건의 구조

설명용 사례 · 실제 API 형식 아님

모델에 허용된 입력

“내가 지나갈 때까지 발판을 밟아 줘.”

AI용 발판 A가 관측됨. 사람이 아직 통과하지 않음.

확정 지시 + 공개 상태
+ 관련 대화·기억 + 실제 실행 결과

→

사람이 검토할 정답

발판 유지

이동 후 멈추는 것만으로는 지시를 충족하지 못합니다.

공개 입력으로 정당한 선택인지 검토

① 행동의 지속 조건을 구분하는 사례

성공·실패를 함께 수집

실패 출력 자체를 정답으로 학습하지 않고, 새 관측에 맞는 선택을 검토합니다.

대기·질문도 포함

이동만 많이 학습하면 기다리거나 물어봐야 하는 상황을 놓칠 수 있습니다.

데이터를 버전으로 보관

모델이 받은 입력, 선택 근거, 계약 버전, 검토 상태를 함께 남깁니다.

한 상태에 여러 유효한 행동이 있을 수 있습니다. 보이지 않는 정보를 알아야만 맞힐 수 있는 답을 강제로 정답으로 만들지 않습니다.

문장만 바꾸는 평가로는 해법 암기를 구분하기 어렵습니다.

먼저 맵·관계·플레이 기록을 분리

그다음 문장 변형·데이터 증강

학습

A 문 1 B

행동 의미를 연습

이동·유지·대기·질문 사례로 모델을 학습합니다.

개발

D C 문 2

설정을 고름

프롬프트·학습 설정·호출 조건을 조정합니다.

시험

문 3 F E

마지막에 평가

새 배치·관계 조합에서 조건을 고정해 비교합니다.

같은 퍼즐의 해법이 학습에 들어가면?

시험 때 관계 정보를 숨겨도 모델이 외워서 맞힐 수 있습니다. 학습 가중치에 들어간 정보도 사전 지식입니다.

어떻게 점검할까?

색·객체 ID·후보 순서를 바꾸고, 같은 플레이의 인접 상태가 학습과 시험에 섞이지 않도록 합니다.

그림은 분리 원리의 예시입니다. 데이터 수와 분할 비율은 실제 수집 범위와 예비 실험 뒤 정합니다.

학습의 효과와 재판단의 효과를 따로 봅니다.

동일한 기반 모델로 만드는 2 × 2 비교

실험 설계 · 결과 없음

	고정 주기 재판단	상황 기반 재판단
학습 전	**A** 기준 조건	**B** 호출 방식 변경
LoRA 적용	**C** 학습만 적용	**D** 두 요소 적용

A–B·C–D는 호출 방식, A–C·B–D는 학습 효과를 비교합니다.

비교할 때 고정할 조건

- ✓ 기반 모델·정밀도·서빙 장비
- ✓ 관측 범위·입력 형식·프롬프트
- ✓ 맵·초기 상태·상황 변화의 종류
- ✓ 파트너 조건·행동 목록·실행 검사
- ✓ 최대 시간·호출 예산과 종료 기준

고정 주기는 여러 간격을 비교합니다. 짧은 프롬프트나 더 작은 모델로 바꾸는 실험은 별도 조건으로 둡니다.

실험 순서

반복 가능한 파트너로 예비 검증한 뒤, 사람과의 협동에서는 순서 효과와 참가자 차이를 고려합니다.

잘 해결했는지, 얼마나 빨랐는지, 얼마나 호출했는지 함께 봅니다.

✓

협동 성공률

전체 시도 중 실제 과제 완료 비율

실패·시간 초과도 분모에 포함

엔진의 접촉·문·통과 이벤트로 판정

◷

반응 시간

상황 변화부터 올바른 새 행동 시작까지

중앙값과 느린 구간을 함께 보고

표본 수와 조건을 명시해 p95 해석

↗

호출 비용

과제당 호출 수·입출력 토큰·실행 시간

성공률과 함께 비교

호출이 적어도 과제에 실패하면 구분

상황 변화 → 관측 갱신 → 모델 판단 → 명령 접수 → 실제 동작 시작

왜 실패했는지도 남깁니다.
잘못된 대상 선택, 불필요한 재호출, 오래된 응답 차단, 순서 위반, 실행 실패와 복구 과정을 분류합니다.

속도는 실제로 측정합니다.
LoRA의 학습 효율과 게임의 응답 속도는 다른 문제입니다. 1초 이하 응답은 목표 의견이며 달성 결과가 아닙니다.

실측 전이므로 성능 그래프나 개선율은 제시하지 않습니다. 효과가 없거나 비용이 증가한 조건도 결과로 보고합니다.

연결 기반은 만들었습니다.
이제 실제 통합과 비교를 진행합니다.

구현·모의 검증

요청에서 실행 결과까지 연결

조정기와 5개 계약, 교체 가능한 관측·모델·실행 어댑터, 실행 직전 검사와 결과 관리를 구현했습니다.

관측 | 판단 입력 | 판단 응답 | 실행 명령 | 실행 결과

기존 모의 왕복·Unreal 출력 회귀 근거. 실제 팀원 퍼즐·모델 통합과 구분합니다.

연구 구현·측정 예정

상황 기반 재판단과 학습

실제 퍼즐·행동 모델 연결, 사건 분류와 반복 판단, 검토된 학습 데이터, LoRA 적용, 네 조건 비교가 남아 있습니다.

영호 조정기·대화·호출 제어·로그

경민 퍼즐·관측·실제 상태 판정

준영 행동 AI·학습·실행 연결

기존 담당을 바탕으로 한 추가 작업 배분안이며 세부 일정은 팀에서 조정합니다.

1 **실제 연결**
10/8 팀 목표
10/9 진행 보고 안내

2 **기준 성능·데이터**
오류와 지연 측정
공개 입력·라벨 검토

3 **학습·호출 정책**
같은 모델에 적용
예비 실험으로 설정 고정

4 **비교와 실패 분석**
새 퍼즐·동일 조건
성과와 한계 함께 보고

새로운 게임 하나와 함께, 다시 검증할 수 있는 결과를 남깁니다.

01

협동 평가 환경

관측 범위·행동·성공 조건·상황 변화를 명세한 퍼즐과 채점기

02

호출 정책과 실행 기록

같은 모델에서 재판단 시점을 바꿔 비교할 조정기와 로그

03

학습 데이터와 모델 기록

검토된 사례, 분리 기준, 기반 모델·LoRA 설정·배포 버전

04

성능과 실패 분석

네 조건의 성공률·지연·호출량과 효과가 없었던 상황까지 정리

파인튜닝하면 무조건 빨라지나요?

아닙니다. 판단 정확도와 재호출이 개선될 가능성을 측정합니다. 모델 크기나 출력 길이를 바꾼 효과는 별도로 봅니다.

정답 관계를 외우는 것 아닌가요?

맵·관계·플레이 단위로 데이터를 나누고, 새로운 관계 조합에서 평가합니다.

이미 있는 연구를 반복하는 것 아닌가요?

알려진 구조와 학습법을 출발점으로 삼아, 명세한 협동 조건에서 두 개선의 효과와 비용을 분리해 검증합니다. 기법 자체의 신규성은 추가 검토 대상입니다.

이번 학기 범위는 어디까지인가요?

기존 퍼즐과 행동 목록에서 시작합니다. 영상 인식·임의 코드 이해·온라인 학습은 핵심 실험에 동시에 넣지 않습니다.

선행연구와 우리 적용 범위를 함께 확인할 수 있습니다.

DPT-Agent ↗ ACL 2025

실시간 협동 구조. 빠른 실행·비동기 추론의 원리를 참고하며 전체 재현은 별도다.

LoRA ↗ 2021

기반 가중치를 고정하는 과제 적응 방법. 게임 속도 향상의 직접 근거는 아니다.

ADAPT ↗ ACL 2026

객체 상태에 따라 달라지는 행동 가능 조건. 영상 관측과 우리의 JSON 입력은 구분한다.

PARTNR ↗ ICLR 2025

역할·공간·순서 제약이 있는 협동 과제와 관측 조건 설계에 참고한다.

Collab-Overcooked ↗ EMNLP 2025

도움 요청과 응답, 과제 진행에 기여하는 협동 과정의 평가에 참고한다.

CoCoBench ↗ arXiv 2026.08

역할·순서·자원·인계별 진단에 참고한다. 사전 공개이며 학회 게재는 확인하지 못했다.

QLoRA ↗ 2023

4비트 기반 모델을 이용한 학습 메모리 절약 방법. 장비 조건을 확인한 뒤 검토한다.

PEFT LoRA 문서 ↗ v0.21.0

어댑터 적용·병합·배포 조건 확인용. 실제 모델과 서빙 버전에서 다시 검증한다.

원문 확인 기준: 2026.10.02. 이 페이지의 퍼즐·시간선·학습 그림은 자체 제작한 설명용 도식이며 논문 실험이나 우리 성능을 재현한 결과가 아닙니다.

처음부터 다시 보기 ↑ 정적인 PDF로 읽기 ↓