

Decision-Oriented PBL Repository 평가 기준 (교수·멘토 관점)

✖ 먼저, 평가하지 않는 것들

교수·멘토는 다음은 거의 보지 않습니다.

- 코드가 얼마나 멋진지
- 최종 결과가 성공했는지
- 하드웨어가 비싼지
- AI를 얼마나 많이 썼는지
- 커밋 메시지가 화려한지

☞ 이 레포의 목적과 무관합니다.

실제로 보는 핵심 기준 7가지

① 문제를 “정답”이 아니라 질문으로 정의했는가

좋은 신호

- “왜 직진 명령인데 한쪽으로 휘는가?”
- “이 현상이 제어 문제인지 하드웨어 문제인지?”

나쁜 신호

- “직진이 안 된다 → 해결해야 한다”

☞ 문제를 고쳐야 할 대상이 아니라

☞ 이해해야 할 현상으로 다루는가를 봅니다.

② Assumption이 명확한가

교수는 이 문장을 가장 먼저 봅니다.

“나는 지금 무엇을 원인이라고 가정하고 있는가?”

좋은 가정

- “좌우 모터의 PWM 응답 차이가 존재할 수 있다”

나쁜 가정

- “뭔가 이상한 것 같다”
- “아마 코드 문제일 것 같다”

☞ 가정은 틀려도 괜찮지만, 흐리면 안 됩니다.

③ Observation이 해석 없이 기록되었는가

멘토는 여기서 학생의 훈련 상태를 바로 압니다.

좋은 관측

- “PWM 80 적용 시 항상 왼쪽으로 휘어짐”
- “배터리 교체 후에도 동일 현상”

나쁜 관측

- “모터가 느린 것 같다”
- “불안정해 보인다”

☞ 관측은 판단이 섞이지 않은 사실이어야 합니다.

④ Interpretation에서 가설을 여러 개 세웠는가

이 항목은 사고의 폭을 봅니다.

좋은 해석

- 가설 A / B / C를 나열
- 왜 A를 먼저 검증하려는지 설명

나쁜 해석

- “원인은 A다” (단정)
- “AI가 이렇게 말함”

☞ 가능성은 열어두는 능력을 평가합니다.

⑤ Next Test가 단일 변수 실험인가

교수는 여기서 실험 설계 능력을 봅니다.

좋은 Next Test

- “좌측 PWM만 -5 조정”
- “배터리·바닥·코드 고정”

나쁜 Next Test

- “여러 값을 바꿔본다”
- “다음엔 PID를 넣어본다”

☞ 한 번에 하나만 바꾸는가가 핵심입니다.

⑥ Episode가 시간 순서로 쌓여 있는가

이건 레포 전체를 보고 판단합니다.

좋은 레포

- Episode 001 → 구조 연습
- Episode 002 → 실제 문제 시작
- Episode 003 → 이전 결과를 이어받음

나쁜 레포

- Episode를 계속 수정
- 번호가 바뀌거나 덮어씀

☞ 사고의 연속성이 보이면 신뢰도가 급상승합니다.

⑦ “틀린 판단”이 삭제되지 않고 남아 있는가

이건 최고 수준의 평가 포인트입니다.

교수가 좋아하는 흔적

- “이 가설은 틀렸음”
- “이 판단은 다음 Episode에서 폐기됨”

☞ 실패 기록이 남아 있으면

☞ 실제 연구·현업 적용력이 높다고 판단합니다.

□ 교수·멘토의 최종 한 줄 평가 기준

이 학생은 문제를 해결하려는가,
아니면 문제를 이해하려는가?

□ Decision Log PBL Repository

-- 평가 기준 가이드 (교수 / 멘토용) --

평가 원칙 한 문장

이 레포는 “무엇을 만들었는가”가 아니라

**“어떻게 판단하며 다음 행동을 설계했는가”를 평가한다.

1. 평가 프레임 개요

평가 대상: LOG.md + Episode 흐름

평가 단위: Episode 묶음 전체

비교 기준: 결과물 ✗ / 사고 구조 ✓

☞ 정답·성공 여부는 감점 요인이 아니다.

2. 핵심 평가 항목 (총 6개)

① 문제 정의 수준 (Problem Framing)

수준	관찰 포인트
○ 낮음	문제가 모호하거나 “안 된다” 수준
◐ 보통	현상을 인식했으나 질문이 흐림
● 높음	현상을 질문 형태로 명확히 정의

멘토 체크 질문

- 이 학생은 문제를 고치려 하는가, 이해하려 하는가?

② 가정 명시 능력 (Assumption Clarity)

수준	관찰 포인트
○ 낮음	가정이 드러나지 않음
◐ 보통	암묵적 가정만 존재
● 높음	“나는 X라고 가정한다”가 명시됨

► 가정은 틀려도 괜찮다.

흐리게 쓰는 것이 문제다.

③ 관측 기록의 객관성 (Observation Quality)

수준	관찰 포인트
○ 낮음	감정·평가가 섞인 서술
◐ 보통	일부 사실, 일부 해석
● 높음	해석 없는 사실 기록

좋은 관측 예

- 수치 / 반복성 / 조건 고정 여부가 드러남

④ 해석의 폭과 태도 (Interpretation Breadth)

수준	관찰 포인트
○ 낮음	단일 원인 단정
◐ 보통	복수 가설 있으나 비교 부족
● 높음	2~3개 가설 + 우선순위 설명

► AI가 한 말 그대로 옮긴 흔적은 감점 요인

⑤ 실험 설계 능력 (Next Test Design)

수준	관찰 포인트
○ 낮음	여러 변수 동시 변경
◑ 보통	실험 의도는 있으나 통제 부족
● 높음	단일 변수 + 명확한 검증 목적

이 항목은 엔지니어링 감각을 가장 잘 보여준다.

⑥ 사고의 연속성 (Episode Continuity)

수준	관찰 포인트
○ 낮음	Episode 수정·삭제 흔적
◑ 보통	흐름은 있으나 연결 약함
● 높음	이전 Episode를 명시적으로 참조

▶ 과거 판단을 “틀렸다고 남겨두는 용기”는 최고점 요소

3. 종합 평가 레벨 (권장)

레벨	의미
● A	연구·현업 적용 가능 사고 구조
◑ B	지도하면 빠르게 성장 가능
○ C	결과 중심 사고에 머물러 있음

▶ A 레벨은 결과가 실패여도 충분히 받을 수 있다.

4. 멘토 코멘트 예시 (바로 사용 가능)

- “가설을 세우는 태도가 좋고, 단일 변수 실험 설계가 명확함”
- “관측과 해석이 섞이는 경향이 있어 분리가 필요함”
- “이전 판단을 수정하지 않고 다음 Episode로 넘긴 점이 인상적임”
- “결과보다 사고 과정이 잘 드러나는 레포”

5. 이 평가 기준의 교육적 의미

- 점수 경쟁 ✕
- 결과 과시 ✕
- 사고 구조 훈련 ✓

☞ AI 시대에 가장 희소한 능력

= 판단을 기록하고 다음 행동을 설계하는 능력

6. 한 줄 요약 (평가자용)

이 레포는

“문제를 맞힌 학생”이 아니라

“문제를 다루는 법을 아는 학생”을 선별하기 위한 도구다.