

AI 에이전트 교육과정 · 모듈 9 부속자료

전부 맡기지도, 전부 막지도 않는 권한 모드 설계 체크리스트

맡길 것과 확인할 것은,
위험도가 정합니다

사용법 안내

이 체크리스트는 AI 에이전트에게 어떤 업무를 자동으로 맡기고, 어떤 업무는 사람이 직접 승인할지 정할 때 작성합니다. **현재 규칙 기준선 → 업무·위험도 분류 → 근거 검증 → 권한 모드 판정 → 예외·중단 조건 → 절차로 남기기** 순서로 한 번 채우고, 실제 업무에 적용하면서 버전을 올려가면 됩니다.

① 현재 규칙 기준선

② 위험도 분류

③ 근거 검증

④ 권한 모드 판정

⑤ 예외·중단 조건

⑥ 절차로 남기기

A 현재 승인 규칙 기준선

지금 AI 업무에 적용 중인 승인 규칙

이 규칙의 근거는 ☐ 실제 데이터 ☐ 타사 사례 확인 ☐ 전해 들은 이야기(소문) ☐ 직접 겪은 경험

이 규칙으로 인한 병목 징후(있다면)

B 업무 목록과 위험도 분류

낮은 위험 — 자동 처리 후보

높은 위험 — 사람 확인 필수

이 구분의 기준(되돌리기 어려운 정도, 외부 노출 여부, 금전 발생 여부 등)

C 근거 검증 — 소문·정보 출처 확인

지금 규칙의 근거가 된 이야기·정보

출처

어떤 조건에서 벌어진 일이었는지

검증 상태 ☐ 검증됨 ☐ 일부 검증 ☐ 미검증

D 권한 모드로 판정

업무 목록과 위험도를 대조한 내용

☐ 자동 처리 ☐ 조건부 자동(사후 확인) ☐ 수동 승인 ☐ 전체 위임 금지(항상 사람) ☐ 추가 검토 필요

E 예외·중단 조건

민감 정보 포함 시 처리 방식

한도 초과·반복 실행 시 처리 방식

이 기준을 팀과 공유했습니까 ☐ 예 ☐ 아니오

F 절차로 남기기

승인 기준 문서화 요약

저장 위치 / 파일명

재검토 주기

기준 버전 v0.1 → 📁 유지 📁 수정 후 v0.2

J 판단 — 우리 조직에 필요한 권한 설계 기준

위 A~E를 근거로, 우리 조직은 무엇을 자동으로 두고 무엇을 사람이 잡아야 합니까?

A 실행 — 권한 기준 확정

확정 기준 요약

공지 방법

다음 재점검일

작성 예시

참고 — A사 대표, "AI 실행 권한 모드 설계" 사례

현재 규칙 기준선: "AI 관련 작업은 전부 대표 승인" 전사 공지, 근거는 동종업계에서 들은 미검증 이야기

위험도 분류: 낮음 — 내부 문서 작성·정리 / 높음 — 결제·외부 전송·대량 파일 변경

근거 검증: "AI가 마음대로 크레딧을 구매했다"는 이야기는 실제로는 한도 없이 반복 실행되다 사용량이 폭증한 사례 — 모든 상황으로 일반화할 수 없었음

권한 모드 판정: 낮은 위험은 자동 처리, 높은 위험은 수동 승인으로 구분

예외·중단 조건: 민감 자료가 포함되면 자동 분류와 무관하게 즉시 멈추고 사람이 확인

절차로 남기기: "AI 업무 승인 기준_v0.1"로 저장, 다음 달 같은 업무에 다시 적용해보고 재점검

QEJA 요약

Q

어떤 업무가 위험하고, 어떤 업무는 자동으로 뒤도 되는가? 지금 규칙은 근거가 있는가?

E

소문의 출처는 확인됐는가? 실제 사례의 발생 조건은 무엇이었는가?

J

업무별로 자동 승인·수동 승인·전체 위임 중 어디에 배정할 것인가?

A

위험도 분류 → 승인 모드 배정 → 문서화 → 팀 공유 → 재점검 → 버전 갱신