정보관리기술사 데일리 다이제스트

2026-10-31 · 국내/국제/공공 12대 도메인 매핑

1. 오늘의 브리핑

축	도메인	헤드라인	키워드
국내	AI IM DS 인공지능	**모델 라우팅 기반 LLM 추론 비용 절감** 작업별 모델 선택 비용 최적화 구조 · itworld.co.kr/article/4242263/%eb%aa%a8%eb%8d%b8-%eb%9d%bc%ec%9a%b0%ed%8c%85%ec%9c%bc%eb%a1%9c-%ec%b6%94%eb%a1%a0-%eb%b9%84%ec%9a%a9-%ec%a4%84%ec%9d%b8%eb%8b%a4%eb%8d%b0%ec%9d%b4%ed%84%b0%eb%b8%8c%eb%a6%ad%ec%8a%a4%c2%b7.html	모델 라우팅 · 추론 비용 · 프론티어 모델 · sLLM
국제	DS CA AI 디지털서비스	**GPU 신뢰성의 분산시스템 문제화 (Netflix)** 계층 헬스체크·자동 복구 설계 사례 · infoq.com/news/2026/10/netflix-gpu-distributed-systems/	GPU 헬스체크 · 자동 복구 · 내결함성 학습 · 신뢰성 플랫폼
공공·정책	LAW SE 법·규제	**안전조치 의무 위반 2개사 과징금 38.85억** **근거** 「개인정보 보호법」 제29조 (안전조치의무) · 안전조치의무 위반 제재 기준 사례 · pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=D010030000&nttId=12600	안전조치 의무 · 과징금 · 개인정보위 · 접근통제

2. 오늘의 토픽 — LLM 모델 라우팅(Model Routing)

AI IM DS SW | 출제 예상: **1교시**

2-1. 개념 및 정의

요청 난이도·비용·품질 요건에 따라 최적 LLM을 동적으로 선택하는 추론 제어 기법. 프론티어 모델 호출 축소를 통한 추론 비용 절감과 품질 유지의 동시 달성.

2-2. 등장 배경

프론티어 모델 상시 사용에 따른 추론 비용 부담 확대. 데이터브릭스·MS·엔비디아·스노우플레이크의 라우팅 기능 도입 확산.

2-3. 핵심 기술요소

분류	주요 기술요소	역할 및 메커니즘
라우팅 판단	난이도 분류기, 품질 예측기, 규칙 기반 정책	요청별 적합 모델 사전 판정
모델 풀	프론티어 모델, sLLM, 구세대 모델	비용·성능 구간별 후보 계층화
운영 제어	캐스케이드 폴백, 비용 예산, 품질 모니터링	저품질 응답 시 상위 모델 재호출

2-4. 시험·실무 관점

- [IM 연계] 토큰 단가 기반 AI 투자 대비 효과 관리
- [DS 연계] 게이트웨이 패턴 기반 다중 모델 서빙 구조
- [SW 연계] 평가셋 기반 라우팅 품질 회귀 테스트

답안 필수 키워드 — 모델 라우팅 / 캐스케이드 / sLLM / 추론 비용 / LLM 게이트웨이