국가과학기술연구회 NAIS 연구직 채용 직무기술서

<table>
<tr><td>채용직종</td><td colspan="2">연구직-2 플랫폼</td><td>채용인원</td><td>2명</td></tr>
<tr><td>모집 분야</td><td colspan="4">LLM 사후학습·학습데이터 구축</td></tr>
<tr><td>수행업무</td><td colspan="4">과학기술 특화 모델 사전·사후학습용 학습데이터 설계·구축 및 사후학습 수행·평가</td></tr>
<tr><td>대분류</td><td colspan="2">중분류</td><td>소분류</td><td>세분류</td></tr>
<tr><td colspan="5">NCS 미개발분야로 자체 직무분석을 통해 도출한 내용임</td></tr>
<tr><td>기관소개</td><td colspan="4">「국가과학AI연구센터」는 국가과학기술연구회 내부조직으로 과학기술 분야의 인공지능 활용 촉진 및 인공지능 기반 연구기관 간의 협업 생태계 조성을 위한 연구개발·지원 역할을 수행함</td></tr>
<tr><td>전형절차</td><td colspan="4">서류전형 → 전공면접(인성검사 포함) → 종합면접 → 임용</td></tr>
<tr><td>직무수행 내용</td><td colspan="4">· 과학기술 특화 모델 학습을 위한 Pre-training 및 post-training 단계별 특화 학습데이터 설계
· 과학기술 도메인·업무 특성을 반영한 데이터 구성 전략 수립
- 데이터 유형 선정, 데이터 혼합 비율 설계, 품질 필터링 기준 정의, 답변 기준 설계
· 학습데이터 자동 생성·정제 파이프라인 구축·운영
· 구축 데이터의 학습 효과 검증 및 도메인별 데이터셋 관리
· 사후학습(SFT·선호 최적화·보상 모델 학습 등) 파이프라인 실행·운영 및 학습 결과 평가·개선 — 정렬 목표·평가 기준은 AI과학자팀(과학 추론 모델·정렬·평가 연구)과 협업</td></tr>
<tr><td rowspan="2">채용예정직급</td><td colspan="4">원 또는 선임*</td></tr>
<tr><td colspan="4">*[선임연구원 임용자격]
· 대학이상의 과정을 이수한 자로서 국가기관 또는 연구기관에서 7년이상**의 연구경력을 가진 자
· 박사학위 소지자로서 국가기관, 대학 또는 연구기관에서 3년이상**의 연구경력을 가진자
· 상기 각 요건과 동등한 자격이 있다고 인정되는 자
** 직급결정 시 경력기간은 내부 경력산정기준에 따름</td></tr>
<tr><td>근무지</td><td colspan="4">대전광역시 또는 업무에 필요한 장소(세종 등)</td></tr>
<tr><td rowspan="2">일반요건</td><td>연령, 성별</td><td colspan="3">· 연령 및 성별 무관</td></tr>
<tr><td>기타</td><td colspan="3">· 「국가공무원법」제33조에 따른 결격사유가 없는 자
· 「병역법」제76조에 따른 병역의무 불이행자에 해당하지 않는 자
· 채용 신체검사 기준에 적합한 자
· 국가연구개발사업 참여 제한이 없는 자
· 연구회 및 출연연 또는 타 공공기관에서 최근 5년 이내 부정한 방법으로 채용된 사실이 적발되어 채용이 취소된 사실이 없는 자
· 「부패방지 및 국민권익위원회의 설치와 운영에 관한 법률」제82조에 따른 취업제한 대상에 해당하지 않는 자
· 임용시기에 근무가 가능한 자
· 대한민국 국적 소지자
· 병역의무 대상자(남성)는 군필자(전역 예정 군인 포함), 면제자</td></tr>
</table>

교육요건	학력	· 석사 이상 학위 소지자 [기졸업자 및 2027년 2월 졸업예정자 포함] ※ 최종 임용되더라도 2027년 2월말까지 학위 취득이 불가한 경우 임용 취소
	관련 전공	인공지능, 컴퓨터공학, 데이터과학 및 이공계 전공 (언어모델 학습·학습데이터 구축 관련 연구실적 또는 실무경력)
필요지식	· LLM 학습 단계별 데이터 요구사항에 관한 지식 및 사전학습 경험 · 데이터 혼합 비율 설계 및 도메인 특화 데이터 커리큘럼 구성에 관한 지식 · 합성 데이터 생성 및 데이터 증강 기법에 관한 지식 · 데이터 품질 필터링, 검증 방법론에 관한 지식 · 과학기술 도메인 데이터(논문·특허·실험 데이터 등)의 특성과 전처리 절차에 관한 지식 · 합성 데이터의 잠재 편향(subliminal bias) 및 안전성에 관한 지식 · AI 모델 서빙·추론 최적화(vLLM 등) 및 배포 방법론에 관한 지식	
필요기술	· Python 기반 대규모 데이터 수집·전처리 파이프라인 개발·운영 기술 · LLM API·오픈소스 모델을 활용한 합성 데이터 자동 생성 및 품질 평가 기술 · 학습 프레임워크(TRL, LLaMA-Factory 등) 활용 및 학습 결과 기반 데이터 개선 기술 · 멀티노드 GPU 분산학습 실행 및 실험 관리·모델 평가 기술 · 학습 실험 결과의 분석·문서화 및 연구 논문·기술보고서 작성 기술	
권장사항	· LLM mid-training/post-training용 특화 데이터셋 설계·구축 실무 경험 · 모델 성능·편향 평가를 위한 벤치마크 데이터셋 설계·구축 경험 · 과학기술 도메인 데이터(논문·실험 데이터 등) 처리 경험 · 오픈소스 데이터셋·모델 공개 실적 · 직무 관련 연구산출물 보유자 (논문, 공개 모델·데이터셋, 공개 저장소·모델 허브, SW, 기술보고서 등)	
직무수행 태도	· 창의적이고 도전적인 연구자세, 객관적인 판단 및 논리적 분석태도 · 투명하고 공정한 업무수행의 청렴성, 문제해결을 위한 적극적 의지 · 조직의 일원으로 구성원과 융화하며 상호 협력하는 자세 · AI 기반 연구환경 변화에 유연하게 적응하고 지속적으로 학습하는 자세	
직업기초 능력	연구직의 경우 필기시험 미전형에 따라 별도의 직업기초능력평가 없음	
참고사이트	· https://nst.re.kr (국가과학기술연구회 홈페이지) · https://nst.fairy.im (국가과학기술연구회 지원서 접수 사이트)	