

분류번호 : 2001070303_23v2

능력단위 명칭 : 인공지능 데이터 확보

능력단위 정의 : 인공지능 데이터 확보란 인공지능 모델에 필요한 데이터 수집 방안을 계획하고, 계획된 방안대로 데이터를 수집하고 증강하는 능력이다.

능 력 단 위 요 소	수 행 준 거
2001070303_23v2.1 인공지능 데이터 확보 계획 수립하기	<p>1.1 인공지능 모델에 필요한 데이터를 정의할 수 있다.</p> <p>1.2 정의된 데이터 확보를 위해 내 외부 확보 가능한 출처(Source)를 파악할 수 있다.</p> <p>1.3 파악한 출처(Source)에서 목표에 맞는 데이터를 수집할 수 있는 출처(Source)를 선정할 수 있다.</p> <p>1.4 선정된 출처(Source)에서 데이터를 수집하기 위한 방법을 정의할 수 있다.</p> <p>1.5 내 외부 데이터의 수집 방법을 종합하여 데이터 통합 확보 계획을 문서화할 수 있다.</p> <p>【지식】</p> <ul style="list-style-type: none">• 인공지능 데이터 이론• 데이터 수집 방법• 인공지능 데이터셋 구축 방법• 인공지능 윤리 가이드북• 인공지능 FATE 가이드북 <p>【기술】</p> <ul style="list-style-type: none">• 인공지능 데이터 학습 기법• IoT 기술(Sensor, Actuator)• 인공지능 인터페이스 기술 <p>【태도】</p> <ul style="list-style-type: none">• 데이터 출처(Source)를 정확하게 파악하려는 노력• 인공지능 윤리기준을 준수하려는 노력
2001070303_23v2.2 인공지능 데이터 수집 환경 구축하기	<p>2.1 데이터 확보 계획에 따라 데이터 출처(Source)의 환경을 파악할 수 있다.</p> <p>2.2 데이터 출처(Source)에서 데이터를 수집할 도구를 설치할 수 있다.</p> <p>2.3 필요한 데이터의 수집, 처리, 저장 단계를 설계할 수 있다.</p> <p>2.4 수집 과정 중 데이터의 오류를 확인하기 위하여 현황과 모니터링 기능을 설계할 수 있다.</p> <p>2.5 고품질의 데이터 확보를 위해 수집 시스템의 관리 기능을 설계할 수 있다.</p> <p>2.6 수집환경의 설계에 따라 데이터 확보 시스템을 구축할 수 있다.</p> <p>【지식】</p> <ul style="list-style-type: none">• 정형·비정형 학습 데이터 환경 구축 방법• 인공지능을 적용할 업무 도메인 지식• 인공지능 데이터 수집 방법 <p>【기술】</p> <ul style="list-style-type: none">• IoT 기술(Sensor, Actuator)• 워드임베딩(Word Embedding)• 언어모형 설계 기술• 음성인식 기술• 기계번역 기술• 영상처리 기술

2001070303_23v2.2 인공지능 데이터 수집 환경 구축하기	【태도】 <ul style="list-style-type: none"> • 내 외부 데이터 수집 항목을 정확하게 식별하려는 노력 • 데이터의 정합성을 정확하게 검증하려는 노력
2001070303_23v2.3 인공지능 내·외부 데이터 수집하기	<p>3.1 구성된 시스템을 활용하여 데이터를 수집할 수 있다. 3.2 수집한 데이터와 원천데이터와의 정합성을 검증할 수 있다. 3.3 검증 결과 부정합 시 데이터를 다시 수집할 수 있다. 3.4 수집된 데이터 출처(Source)와 사용 용도를 기록할 수 있다.</p> 【지식】 <ul style="list-style-type: none"> • 정형·비정형 학습 데이터 환경 구축 방법 • 인공지능에 필요한 내·외부 데이터 수집 방법론 【기술】 <ul style="list-style-type: none"> • 워드임베딩(Word Embedding) • 인공지능 데이터 전처리 기술 • 영상처리 기술 • 음성 처리 기술 • 자연어 처리 기술 【태도】 <ul style="list-style-type: none"> • 데이터 수집 항목을 정확하게 식별하는 노력 • 수집 프로세스 흐름에 대한 주의 깊게 관찰하는 자세
2001070303_23v2.4 인공지능 데이터 증강하기	<p>4.1 확보된 데이터가 부족할 시에는 유사한 데이터 증강 계획을 수립할 수 있다. 4.2 데이터를 증강하기 위한 다양한 기법 중에서 적합한 기법을 선택하고 적용할 수 있다. 4.3 수립된 증강 계획에 따라 수집된 데이터를 필요한 형태로 가공할 수 있다. 4.4 증강된 데이터를 기존 데이터셋에 병합할 수 있다.</p> 【지식】 <ul style="list-style-type: none"> • 인공지능에 필요한 데이터 증강 방법 • 정형·비정형 학습 데이터 환경 구축 방법 • 인공지능에 필요한 내·외부 데이터 수집 방법론 • 생성(Generative) AI 활용 방법론 【기술】 <ul style="list-style-type: none"> • 인공지능 데이터 학습 기법 • 이미지 오그멘테이션(image augmentation) 기법 • 바이오 오그멘테이션(Bio augmentation) 기법 • 대형 언어 모델(LLM) 기술 • 데이터 증강 시의 타당성 검증 노력 【태도】 <ul style="list-style-type: none"> • 데이터 수집에 관한 문제해결을 위한 긍정적 노력 • 심도 깊게 증강 계획을 수립하려는 노력

□ 적용범위 및 작업상황

고려사항

-이 능력단위는 인공지능 학습을 위하여 데이터 확보 계획을 수립하고, 데이터를 수집하고 증강하여 충분한 데이터를 확보하는 업무에 적용한다.

-인공지능 데이터란 인공지능 모델을 학습시키기 위해 사용되는 데이터를 말한다. 이 데이터는 모델이 특정 작업을 수행할 때 필요한 정보를 담고 있으며, 모델이 예측하는 결과에 대한 지도 역할을 한다.

-인공지능 데이터셋은 다양한 형태와 크기로 구성될 수 있다. 일부 데이터셋은 수백만 개 이상의 이미지, 비디오, 텍스트 등을 포함하고 있으며, 다른 데이터셋은 몇 개의 샘플만 포함할 수도 있다. 데이터셋은 학습 데이터와 테스트 데이터로 나뉘며, 일반적으로 학습 데이터는 모델을 학습시키는 데 사용되고, 테스트 데이터는 모델의 성능을 평가하는 데 사용된다.

-내부 데이터(Internal Data)는 기업이나 조직 내부에서 생성되거나 수집한 데이터를 말한다. 예를 들어, 기업의 고객 정보, 판매 기록 등이 내부 데이터의 예시이다. 이러한 내부 데이터는 기업의 내부 시스템에서 추출되거나 수집되며, 기업 내부에서만 사용된다.

-외부 데이터(External Data)는 외부 소스에서 수집한 데이터를 말한다. 예를 들어, 인터넷에서 수집한 이미지, 뉴스 기사, 공공 데이터 등이 외부 데이터의 예시이다. 이러한 외부 데이터는 기업이나 조직 외부에서 수집되며, 기업 내부에서 사용되기도 한다.

-인공지능 데이터 확보 계획 수립은 인공지능 서비스에 필요한 데이터를 정의하고, 내 외부 확보 가능한 출처(Source)를 파악하여 내부 확보 가능한 데이터는 데이터의 존재 형태, 외부 확보 가능한 데이터는 데이터의 존재 형태를 파악하고, 이를 종합하여 데이터 확보 계획을 수립하는 능력이다.

-인공지능 데이터 확보 시스템 구성은 데이터 발생의 근원지를 파악하여 거기에 데이터를 인식할 도구를 설치하고, 인식된 데이터를 확보하기 위해 데이터의 수집, 처리, 저장 단계와 현황 및 모니터링 기능, 그리고 수집 시스템의 총괄 관리 기능을 설계하여 데이터 확보 시스템을 구성하는 능력이다.

-인공지능 내·외부 데이터 수집은 내부 데이터 확보를 위해 ETL(추출, 변환, 로드)을 설계하여 데이터를 확보하거나 수집을 위해 설치한 도구를 활용하여 외부 데이터 확보를 확보하고, 수집한 데이터를 검증하기 위하여 원천데이터와 검증을 수행하며, 검증 결과에 따라 데이터를 다시 수집하는 능력이다.

-인공지능 데이터 증강은 확보된 데이터가 부족할 시에는 유사한 데이터 생성을 계획하고, 수집된 데이터의 특징(Feature)을 파악하여 변형된 특징을 가진 유사 데이터를 생성하며, 수집 데이터를 학습하여 나온 모델에 대해서는 유사 데이터를 더하여 재학습하고, 재학습 모델의 성능이 최대화될 때까지 새로운 유사데이터를 생성하고 재학습하는 것을 반복하는 능력이다.

-인공지능 FATE란 인공지능의 공정성(Fairness), 책임성(Accountability), 투명성(Transparency), 윤리의식(Ethics)을 의미한다.

-인공지능 윤리기준이란 3대 기본원칙에 따라 인간성(Humanity)을 구현하기 위해 인공지능의 개발 및 활용 과정에서 인간의 존엄성 원칙, 사회의 공공선 원칙, 기술의 합목적성 원칙을 지키는 것을 의미한다.

자료 및 관련 서류

- 인공지능 데이터 확보 계획서
- 인공지능 데이터 확보 시스템 설계서
- 인공지능 데이터 수집 품질 관리서
- 인공지능 데이터 증강 계획서

장비 및 도구

- 컴퓨터, 문서작성 프로그램
- 데이터 분석용 소프트웨어 또는 패키지
- 분석 데이터 처리용 프로그래밍 언어 실행환경
- 데이터 시각화 프로그램

재료

- 해당사항 없음

□ 평가지침

권장평가방법

- 평가자는 능력단위 인공지능 데이터 확보의 수행준거에 제시되어 있는 내용을 평가하기 위해 이론과 실기를 나누어 평가하거나 종합적인 결과물의 평가 등 다양한 평가 방법을 사용할 수 있다.
- 평가자는 다음 사항을 평가해야 한다.

권 장 평 가 방 법	평 가 유 형	
	과 정 평 가	결 과 평 가
A.포트폴리오		
B.문제해결 시나리오	V	
C.서술형시험	V	V
D.논술형시험		V
E.사례연구	V	
F.평가자 질문		V
G.평가자 체크리스트	V	V
H.피평가자 체크리스트		
I.일지/저널		
J.역할연기		
K.구두발표	V	
L.작업장평가		
M.기타		

평가지 고려사항

- 수행준거에 제시되어 있는 내용을 성공적으로 수행할 수 있는지를 평가해야 한다.
 - 평가자는 다음 사항을 평가해야 한다.
- 인공지능 데이터 확보 계획의 작성 능력
 - 인공지능 데이터 확보 시스템의 구성 항목
 - 인공지능 내·외부 데이터 수집 절차 및 문제점 해결 능력
 - 인공지능 데이터 증강시키는 능력
 - 인공지능 데이터 확보 관련 문서 작성 능력

□ 관련기초능력

순번	관 련 기 초 능 력	
	주 요 영 역	하 위 영 역
1	의사소통능력	경청 능력, 기초외국어 능력, 문서이해 능력, 문서작성 능력, 의사표현 능력
2	수리능력	기초연산 능력, 기초통계 능력, 도표분석 능력, 도표작성 능력
3	문제해결능력	문제처리 능력, 사고력
4	정보능력	정보처리 능력, 컴퓨터활용 능력
5	기술능력	기술선택 능력, 기술이해 능력, 기술적용 능력

□ 개발·개선 이력

구 분		내 용
직무명칭(능력단위명)		인공지능모델링(인공지능 데이터 확보)
분류번호	기존	2001070303_19v1
	현재	2001070303_23v2
개발·개선연도	현재	2023
	최초(1차)	2019
버전번호		v2
개발·개선기관	현재	
	최초(1차)	정보기술·사업관리 인적자원개발위원회(한국IT비즈니스진흥협회)
향후 보완 연도(예정)		-