

AiNex 프로젝트 AI/AX 용어집 (Glossary)

문서 버전: 1.0

최종 수정일: 2026 년 1 월 7 일

작성자: AiNex Consulting Team

Copyright: © 2025 A3 Security.Co.,Ltd. R&D Center. All rights reserved.

개요

본 용어집은 AiNex AI 컨설팅 프로젝트 수행 시 필수적으로 알아야 할 AI/AX 관련 용어를 집대성한 것입니다. AI 전략 수립부터 기술 구현, 거버넌스, 윤리, MLOps 에 이르기까지 프로젝트 전반에 걸친 핵심 용어들을 체계적으로 정리하였습니다.

용어집 활용 방법

- 컨설턴트: 고객사와의 커뮤니케이션 시 정확한 용어 사용
- 이용자: AI/AX 프로젝트 이해도 향상 및 의사결정 지원
- 개발자: 기술 용어의 정확한 의미 파악 및 구현 가이드

한글 용어 (가나다순)

ㄱ

가명화 (Pseudonymization)

정의: 개인 식별자를 다른 값으로 대체하여 개인을 직접 식별할 수 없도록 만드는 비식별화 기법

활용: 개인정보 보호법 준수를 위한 데이터 전처리

관련 용어: 비식별화, 데이터 프라이버시, 마스킹

예시: 고객 이름을 임의의 코드로 대체 (홍길동 → USER_12345)

가치-실행 매핑 (Value-Feasibility Mapping)

정의: AI 도입 과제를 비즈니스 가치와 실행 용이성 두 차원으로 평가하여 우선순위를 결정하는 방법론

활용: AI 프로젝트 우선순위 결정

관련 용어: 2x2 매트릭스, Quick Win, 전략적 과제

예시: 높은 가치 + 높은 용이성 = Quick Win 과제로 분류

거버넌스 (Governance)

정의: AI 시스템의 개발, 배포, 운영 전반에 걸친 규칙, 정책, 프로세스, 책임 체계

활용: AI 시스템의 책임 있는 사용 보장

관련 용어: 거버넌스 프레임워크, 정책, 규제 준수

예시: AI 거버넌스 위원회 설립 및 운영

경영진 스폰서십 (Executive Sponsorship)

정의: 최고경영진의 AI 추진 의지와 지원 수준

활용: AI 프로젝트 성공을 위한 조직적 지원 확보

관련 용어: 전략적 정렬, 리더십, 변화 관리

예시: CEO의 AI 전환 전략 선언 및 예산 승인

공정성 (Fairness)

정의: AI 시스템이 특정 그룹이나 특성을 가진 사람들에게 차별적으로 작동하지 않는 특성

활용: 편향성 방지 및 윤리적 AI 구현

관련 용어: 편향성, 차별 방지, 윤리

예시: 채용 AI가 성별에 따라 다른 결과를 내지 않도록 보장

공정성 지표 (Fairness Metrics)

정의: AI 모델의 공정성을 정량적으로 측정하는 지표

활용: 편향성 측정 및 모니터링

관련 용어: Demographic Parity, Equalized Odds, 편향성 테스트

예시: 성별별 모델 정확도 차이 측정

과적합 (Overfitting)

정의: 모델이 학습 데이터에 지나치게 적합되어 새로운 데이터에 대한 일반화 성능이 떨어지는 현상

활용: 모델 성능 향상 및 일반화 능력 개선

관련 용어: 일반화, 검증 데이터, 정규화

예시: 학습 정확도 99%, 테스트 정확도 70%인 경우

L

내부 감사 (Internal Audit)

정의: 조직 내부에서 AI 거버넌스 체계와 프로세스의 준수 여부를 정기적으로 검토하는 활동

활용: 거버넌스 준수 확인 및 개선점 도출

관련 용어: 규제 준수, 감사 추적, 준수 관리

예시: 분기별 AI 모델 배포 절차 준수 여부 감사

데이터 거버넌스 (Data Governance)

정의: 데이터의 수집, 저장, 사용, 폐기 전 과정에 대한 정책과 절차

활용: 데이터 품질 및 보안 관리

관련 용어: 데이터 품질, 메타데이터 관리, 데이터 정책

예시: 데이터 접근 권한 정책 수립 및 시행

데이터 드리프트 (Data Drift)

정의: 시간이 지남에 따라 입력 데이터의 통계적 특성이 변화하는 현상

활용: 모델 성능 저하 감지 및 재학습 필요성 판단

관련 용어: 개념 드리프트, 모델 모니터링, 재학습

예시: 고객 행동 패턴 변화로 인한 추천 모델 성능 저하

데이터 레이크 (Data Lake)

정의: 구조화되지 않은 원시 데이터를 그대로 저장하는 대용량 저장소

활용: 다양한 형태의 데이터를 원본 그대로 보존

관련 용어: 데이터 웨어하우스, 빅데이터, 데이터 파이프라인

예시: AWS S3 에 로그, 이미지, 센서 데이터 등을 원본 형식으로 저장

데이터 리니지 (Data Lineage)

정의: 데이터의 출처, 변환 과정, 사용 이력을 추적하는 체계

활용: 데이터 신뢰성 확보 및 문제 원인 파악

관련 용어: 데이터 추적, 메타데이터, 감사 추적

예시: 보고서 데이터가 어느 원천 시스템에서 어떤 변환을 거쳐 생성되었는지 추적

데이터 웨어하우스 (Data Warehouse)

정의: 구조화된 데이터를 분석 목적으로 최적화하여 저장하는 저장소

활용: 빠른 쿼리 성능 및 비즈니스 인텔리전스

관련 용어: ETL, OLAP, 데이터 마트

예시: AWS Redshift 에서 매출 데이터 분석 쿼리 실행

데이터 증강 (Data Augmentation)

정의: 기존 데이터를 변형하거나 합성하여 학습 데이터를 늘리는 기법

활용: 데이터 부족 문제 해결 및 모델 일반화 성능 향상

관련 용어: 오버샘플링, SMOTE, 합성 데이터

예시: 이미지 회전, 반전, 밝기 조절을 통한 학습 데이터 증강

데이터 파이프라인 (Data Pipeline)

정의: 데이터를 수집, 변환, 저장하는 일련의 자동화된 프로세스

활용: 데이터 처리 자동화 및 효율성 향상

관련 용어: ETL, ELT, 데이터 플로우

예시: Apache Airflow 를 사용한 일일 데이터 수집 및 전처리 자동화

데이터 품질 (Data Quality)

정의: 데이터의 정확성, 완전성, 일관성, 시의성 등의 특성

활용: 모델 성능 및 분석 결과 신뢰성 확보

관련 용어: 데이터 검증, 이상치, 결측치

예시: 데이터 품질 점수 1~5 점 척도로 평가

데이터 프라이버시 (Data Privacy)

정의: 개인정보를 포함한 민감한 데이터를 보호하기 위한 체계

활용: 개인정보 보호법 및 GDPR 준수

관련 용어: 개인정보 보호, 비식별화, 동의 관리

예시: 개인정보 비식별화 후 AI 모델 학습 데이터로 사용

데이터 시트 (Data Sheet)

정의: 데이터셋의 특성, 출처, 수집 방법, 제한사항 등을 문서화한 문서

활용: 데이터 투명성 확보 및 적절한 사용 유도

관련 용어: 모델 카드, 데이터 문서화, 메타데이터

예시: 학습 데이터셋의 수집 기간, 샘플 수, 레이블링 방법 등을 문서화

딥러닝 (Deep Learning)

정의: 다층 신경망을 사용하여 복잡한 패턴을 학습하는 머신러닝 기법

활용: 이미지 인식, 자연어 처리, 음성 인식 등

관련 용어: 신경망, CNN, RNN, Transformer

예시: ResNet 을 사용한 이미지 분류 모델 개발

르

로드맵 (Roadmap)

정의: AI 도입 전략의 단계별 실행 계획을 시간 순서대로 나타낸 계획서

활용: AI 프로젝트의 체계적 추진 및 진행 상황 관리

관련 용어: 전략, 마일스톤, 일정

예시: 3 년간의 AI 플랫폼 구축 로드맵 (1 년차: 기반 구축, 2 년차: 확산, 3 년차: 최적화)

롤백 (Rollback)

정의: 배포된 모델이나 시스템을 이전 버전으로 되돌리는 작업

활용: 문제 발생 시 빠른 복구

관련 용어: 배포, 자동 롤백, 버전 관리

예시: 새 모델 배포 후 성능 저하 발견 시 자동으로 이전 버전으로 복구

리스크 관리 (Risk Management)

정의: AI 시스템의 위험을 식별, 평가, 처리하는 체계적 프로세스

활용: AI 프로젝트의 위험 최소화

관련 용어: 위험 평가, 리스크 등록부, 완화 전략

예시: AI 모델 편향 리스크 식별 및 완화 계획 수립

□

마이크로서비스 (Microservices)

정의: 애플리케이션을 작고 독립적인 서비스로 나누어 개발하는 아키텍처 패턴

활용: 확장성, 유지보수성 향상

관련 용어: 컨테이너, API, 서비스 오케스트레이션

예시: 모델 서빙, 데이터 수집, 전처리를 각각 독립적인 서비스로 구현

모델 드리프트 (Model Drift)

정의: 시간이 지남에 따라 모델의 성능이 저하되는 현상

활용: 모델 재학습 필요성 판단

관련 용어: 데이터 드리프트, 개념 드리프트, 모델 모니터링

예시: 고객 이탈 예측 모델의 정확도가 90%에서 75%로 하락

모델 레지스트리 (Model Registry)

정의: 학습된 모델의 버전, 메타데이터, 성능 지표를 중앙에서 관리하는 시스템

활용: 모델 버전 관리 및 추적

관련 용어: MLflow, 모델 버전 관리, 모델 스테이징

예시: MLflow Model Registry 에 모델 버전별로 성능 메트릭과 학습 파라미터 저장

모델 카드 (Model Card)

정의: AI 모델의 특성, 성능, 제한사항, 사용 방법 등을 문서화하는 표준 문서

활용: 모델 투명성 확보 및 적절한 사용 유도

관련 용어: 모델 문서화, 투명성, 설명 가능성

예시: 모델 이름, 알고리즘, 성능 지표, 제한사항, 편향 분석 결과 등을 문서화

모니터링 (Monitoring)

정의: AI 시스템의 성능, 상태, 메트릭을 지속적으로 관찰하는 활동

활용: 문제 조기 발견 및 대응

관련 용어: 로깅, 알림, 대시보드

예시: Prometheus 와 Grafana 를 사용한 모델 성능 실시간 모니터링

머신러닝 (Machine Learning)

정의: 데이터로부터 패턴을 학습하여 예측이나 결정을 수행하는 AI 기술

활용: 분류, 회귀, 군집화, 추천 등 다양한 문제 해결

관련 용어: 지도 학습, 비지도 학습, 강화 학습

예시: 고객 이탈 예측, 상품 추천, 이미지 분류

ㅂ

배치 처리 (Batch Processing)

정의: 일정 시간 간격으로 대량의 데이터를 한 번에 처리하는 방식

활용: 정기적인 대량 데이터 처리

관련 용어: 실시간 스트리밍, ETL, 스케줄링

예시: 매일 밤 전날 거래 데이터를 일괄 처리하여 보고서 생성

배포 (Deployment)

정의: 학습된 모델을 실제 운영 환경에 적용하는 과정

활용: 모델을 실제 서비스에 적용

관련 용어: CI/CD, 모델 서빙, 배포 전략

예시: 학습된 추천 모델을 프로덕션 서버에 배포하여 실시간 추천 제공

벤치마킹 (Benchmarking)

정의: 동종 업계나 유사 조직과의 비교를 통해 상대적 위치를 파악하는 방법

활용: 조직의 AI 역량 수준 평가

관련 용어: 성숙도 평가, 비교 분석

예시: 동종 업계 AI 도입 수준과 자사 수준 비교

변화 관리 (Change Management)

정의: 조직 변화를 계획하고 실행하며 정착시키는 체계적 접근 방법

활용: AI 도입 시 조직의 저항 최소화 및 변화 수용 촉진

관련 용어: 조직 문화, 변화 관리 전략, 교육

예시: AI 도입에 따른 업무 프로세스 변화에 대한 직원 교육 및 지원

비식별화 (De-identification)

정의: 개인을 식별할 수 있는 정보를 제거하거나 변환하는 과정

활용: 개인정보 보호 및 법규 준수

관련 용어: 가명화, 마스킹, 데이터 프라이버시

예시: 이름, 주민등록번호 등 개인 식별 정보 제거 또는 변환

편향성 (Bias)

정의: AI 모델이 특정 그룹이나 상황에 대해 불공정하거나 차별적인 결과를 내는 현상

활용: 공정성 확보 및 윤리적 AI 구현

관련 용어: 공정성, 차별, 편향 완화

예시: 채용 AI가 특정 성별이나 인종에 대해 부당하게 낮은 점수 부여

편향성 테스트 (Bias Testing)

정의: AI 모델이 특정 그룹에 대해 편향된 결과를 생성하는지 검사하는 테스트

활용: 모델의 공정성 평가

관련 용어: 공정성 지표, 편향 분석, 그룹별 성능 비교

예시: 성별, 연령별로 모델 정확도를 비교하여 편향 여부 확인

人

성숙도 모델 (Maturity Model)

정의: 조직의 역량 수준을 단계별로 평가하는 프레임워크

활용: 현재 수준 파악 및 개선 방향 설정

관련 용어: CMMI, 성숙도 레벨, 역량 평가

예시: AI 성숙도 5 단계 - 초기, 반복, 정의, 관리, 최적화

설명 가능성 (Explainability)

정의: AI 시스템의 의사결정 과정과 결과를 설명할 수 있는 특성

활용: AI 신뢰성 확보 및 투명성 보장

관련 용어: XAI, 해석 가능성, SHAP, LIME

예시: 대출 승인 AI가 특정 고객을 거절한 이유를 구체적으로 설명

센서 데이터 (Sensor Data)

정의: IoT 센서나 측정 장치에서 수집되는 물리적 환경 데이터

활용: 제조업 품질 예측, 예측 정비 등

관련 용어: IoT, 시계열 데이터, 실시간 데이터

예시: 온도, 압력, 진동 센서 데이터를 사용한 설비 고장 예측

스트리밍 처리 (Streaming Processing)

정의: 데이터가 생성되는 즉시 실시간으로 처리하는 방식

활용: 실시간 분석 및 빠른 의사결정

관련 용어: Apache Kafka, 실시간 처리, 이벤트 기반

예시: 실시간 사기 탐지 시스템에서 거래 데이터를 즉시 분석

시계열 데이터 (Time Series Data)

정의: 시간 순서대로 기록된 데이터

활용: 추세 분석, 예측, 이상 탐지

관련 용어: 시계열 예측, LSTM, ARIMA

예시: 주가 데이터, 센서 데이터, 매출 데이터

시나리오 분석 (Scenario Analysis)

정의: 다양한 가정 하에서 여러 시나리오를 생성하고 비교 분석하는 방법

활용: 의사결정 지원 및 리스크 평가

관련 용어: 보수적 시나리오, 균형 시나리오, 공격적 시나리오

예시: AI 투자 규모에 따른 3 가지 시나리오별 ROI 분석

○

윤리 (Ethics)

정의: AI 시스템의 개발과 사용에서 지켜야 할 도덕적 원칙과 가치

활용: 책임 있는 AI 구현

관련 용어: AI 윤리 원칙, 사회적 책임, 인간 중심 AI

예시: AI 윤리 원칙 선언 - 공정성, 투명성, 책임성, 안전성

의사결정 로깅 (Decision Logging)

정의: AI 시스템이 내린 모든 의사결정을 기록하는 체계

활용: 의사결정 추적 및 감사

관련 용어: 감사 추적, 로그 관리, 추적 가능성

예시: AI 가 대출을 거절한 모든 경우에 대해 입력, 출력, 이유를 기록

이상 탐지 (Anomaly Detection)

정의: 정상 패턴에서 벗어난 비정상적인 데이터나 행동을 식별하는 기술

활용: 사기 탐지, 품질 관리, 보안

관련 용어: 이상치, 아웃라이어, 비정상 탐지

예시: 신용카드 사기 거래 탐지, 제조 공정 이상 감지

인공지능 (Artificial Intelligence, AI)

정의: 인간의 지능적 행동을 모방하는 컴퓨터 시스템

활용: 다양한 지능적 작업 자동화

관련 용어: 머신러닝, 딥러닝, 자연어 처리, 컴퓨터 비전

예시: 음성 인식, 이미지 인식, 자연어 이해, 의사결정 지원

ㅈ

자동화 (Automation)

정의: 사람의 개입 없이 작업이 자동으로 수행되도록 하는 것

활용: 효율성 향상 및 인력 절감

관련 용어: RPA, 자동 배포, CI/CD

예시: 모델 재학습 및 배포 과정 자동화

자연어 처리 (Natural Language Processing, NLP)

정의: 인간의 언어를 컴퓨터가 이해하고 처리하는 기술

활용: 텍스트 분석, 챗봇, 번역, 요약

관련 용어: BERT, GPT, Transformer, 토큰화

예시: 고객 리뷰 감정 분석, 챗봇 응답 생성

재현성 (Reproducibility)

정의: 동일한 조건에서 실험을 반복했을 때 동일한 결과를 얻을 수 있는 특성

활용: 연구 신뢰성 확보 및 검증

관련 용어: 버전 관리, 시드 값, 실험 추적

예시: 동일한 하이퍼파라미터와 데이터로 모델을 재학습했을 때 동일한 성능 달성

전략적 정렬 (Strategic Alignment)

정의: AI 전략이 비즈니스 전략과 연계되어 있는 정도

활용: AI 투자의 비즈니스 가치 극대화

관련 용어: 비즈니스 전략, AI 비전, 전략적 목표

예시: 고객 만족도 향상이라는 비즈니스 목표와 AI 추천 시스템 구축 전략의 연계

전이 학습 (Transfer Learning)

정의: 한 도메인에서 학습한 모델을 다른 도메인에 적용하는 기법

활용: 학습 시간 단축 및 소량 데이터로 고성능 달성

관련 용어: 사전 학습 모델, Fine-tuning, ImageNet

예시: ImageNet 으로 사전 학습된 ResNet 을 의료 이미지 분류에 적용

정밀도 (Precision)

정의: 모델이 양성으로 예측한 것 중 실제로 양성인 비율

활용: 분류 모델 성능 평가

관련 용어: 재현율, F1 Score, 혼동 행렬

예시: 스팸 메일로 분류한 100 개 중 실제로 스팸인 것이 90 개이면 정밀도 90%

정형 데이터 (Structured Data)

정의: 행과 열로 구성된 표 형태의 데이터

활용: 데이터베이스 저장 및 분석

관련 용어: 관계형 데이터베이스, CSV, SQL

예시: 고객 정보 테이블 (이름, 나이, 주소, 구매 이력)

준수 관리 (Conformance Management)

정의: 법규, 규제, 표준을 준수하는지 관리하는 체계

활용: 법적 리스크 최소화

관련 용어: 규제 준수, 컴플라이언스, 내부 감사

예시: GDPR 준수를 위한 개인정보 처리 절차 수립 및 감사

大

추론 (Inference)

정의: 학습된 모델을 사용하여 새로운 데이터에 대한 예측을 수행하는 과정

활용: 실제 서비스에서 모델 활용

관련 용어: 예측, 모델 서빙, 추론 시간

예시: 학습된 이미지 분류 모델을 사용하여 새 이미지의 카테고리 예측

추천 시스템 (Recommendation System)

정의: 사용자에게 적합한 항목을 추천하는 AI 시스템

활용: 개인화 서비스 제공

관련 용어: 협업 필터링, 콘텐츠 기반 필터링, 하이브리드

예시: 넷플릭스 영화 추천, 아마존 상품 추천

크

컨테이너 (Container)

정의: 애플리케이션과 실행 환경을 함께 패키징한 경량 가상화 기술

활용: 일관된 실행 환경 제공 및 배포 간소화

관련 용어: Docker, Kubernetes, 컨테이너 오케스트레이션

예시: Docker 이미지로 모델 서빙 애플리케이션 패키징

컴퓨터 비전 (Computer Vision)

정의: 이미지나 비디오를 이해하고 분석하는 AI 기술

활용: 이미지 분류, 객체 탐지, 얼굴 인식

관련 용어: CNN, 이미지 인식, 객체 탐지

예시: 자율주행 차량의 도로 상황 인식

ㅅ

테스트 데이터 (Test Data)

정의: 모델의 최종 성능을 평가하기 위해 사용하는 데이터셋

활용: 모델의 일반화 성능 평가

관련 용어: 학습 데이터, 검증 데이터, 데이터 분할

예시: 전체 데이터의 20%를 테스트 데이터로 분리하여 최종 평가

투명성 (Transparency)

정의: AI 시스템의 작동 방식과 의사결정 근거를 이해할 수 있는 특성

활용: AI 신뢰성 및 책임성 확보

관련 용어: 설명 가능성, 모델 카드, 데이터 시트

예시: 모델의 입력, 출력, 의사결정 과정을 명확히 문서화

ㅇ

파이프라인 (Pipeline)

정의: 데이터 처리나 모델 개발의 일련의 단계를 자동화한 워크플로우

활용: 반복 작업 자동화 및 효율성 향상

관련 용어: MLOps, CI/CD, 워크플로우

예시: 데이터 수집 → 전처리 → 학습 → 평가 → 배포의 자동화된 파이프라인

특성 공학 (Feature Engineering)

정의: 원시 데이터로부터 모델 학습에 유용한 특성을 생성하는 과정

활용: 모델 성능 향상

관련 용어: 특성 선택, 특성 변환, Feature Store

예시: 거래 날짜로부터 요일, 월, 분기 등의 특성 생성

ㅎ

하이브리드 LLM (Hybrid LLM)

정의: 로컬 LLM 과 온라인 LLM 을 함께 사용하는 방식

활용: 보안과 성능의 균형

관련 용어: Ollama, Local LLM, 데이터 보안

예시: 민감한 데이터는 로컬 LLM 으로 처리하고, 일반 데이터는 온라인 LLM 사용

하이퍼파라미터 (Hyperparameter)

정의: 모델 학습 전에 설정하는 파라미터 (학습률, 배치 크기 등)

활용: 모델 성능 최적화

관련 용어: 하이퍼파라미터 튜닝, Grid Search, Optuna

예시: 학습률 0.001, 배치 크기 32, 에포크 100 으로 설정

학습 (Training)

정의: 데이터로부터 모델의 파라미터를 최적화하는 과정

활용: AI 모델 개발

관련 용어: 손실 함수, 옵티마이저, 역전파

예시: 10,000 개의 이미지로 분류 모델을 100 에포크 동안 학습

학습 데이터 (Training Data)

정의: 모델을 학습시키기 위해 사용하는 데이터셋

활용: 모델의 파라미터 최적화

관련 용어: 레이블, 지도 학습, 데이터 증강

예시: 레이블이 있는 10,000 개의 이미지 데이터로 분류 모델 학습

혼동 행렬 (Confusion Matrix)

정의: 분류 모델의 예측 결과를 실제 값과 비교하여 나타낸 표

활용: 모델의 오분류 패턴 분석

관련 용어: 정밀도, 재현율, True Positive, False Positive

예시: 2x2 행렬로 TP, TN, FP, FN 값을 표시

영문 용어 (ABC 순)

A

A/B Testing

정의: 두 가지 버전의 모델을 동시에 운영하여 성능을 비교하는 테스트 방법

활용: 새 모델의 성능 검증

관련 용어: Canary 배포, 트래픽 분할, 실험

예시: 기존 추천 모델과 새 모델을 50:50 으로 트래픽을 나누어 성능 비교

API (Application Programming Interface)

정의: 소프트웨어 간 상호작용을 위한 인터페이스

활용: 시스템 통합 및 서비스 제공

관련 용어: REST API, gRPC, API Gateway

예시: 모델 예측을 HTTP REST API 로 제공

AUC-ROC (Area Under the ROC Curve)

정의: 이진 분류 모델의 성능을 나타내는 지표

활용: 분류 모델의 전반적인 성능 평가

관련 용어: ROC 곡선, TPR, FPR

예시: AUC-ROC 0.92 는 우수한 분류 성능을 의미

Auto-scaling

정의: 부하에 따라 자동으로 리소스를 확장하거나 축소하는 기능

활용: 비용 최적화 및 성능 보장

관련 용어: 탄력성, 클라우드, 동적 스케일링

예시: 트래픽 증가 시 자동으로 서버 인스턴스 추가

AWS (Amazon Web Services)

정의: 아마존이 제공하는 클라우드 컴퓨팅 플랫폼

활용: AI/ML 인프라 구축

관련 용어: S3, EC2, SageMaker, Lambda

예시: AWS S3 에 학습 데이터 저장, SageMaker 로 모델 학습

B

Batch Size

정의: 모델 학습 시 한 번에 처리하는 샘플의 개수

활용: 학습 속도와 메모리 사용량 조절

관련 용어: 에포크, 이터레이션, 미니배치

예시: 배치 크기 32 로 설정하여 학습

BERT (Bidirectional Encoder Representations from Transformers)

정의: 구글이 개발한 양방향 트랜스포머 기반 자연어 처리 모델

활용: 텍스트 분류, 질의응답, 개체명 인식

관련 용어: Transformer, 사전 학습 모델, Fine-tuning

예시: 고객 리뷰 감정 분석에 BERT-base 모델 사용

Bias (편향)

정의: 모델이 특정 방향으로 치우친 예측을 하는 경향

활용: 공정성 확보 및 모델 개선

관련 용어: Fairness, 차별, 편향 완화

예시: 채용 AI 가 특정 성별을 선호하는 편향 발견

Big Data

정의: 기존 방법으로는 처리하기 어려운 대용량 데이터

활용: 대규모 데이터 분석 및 인사이트 도출

관련 용어: 데이터 레이크, Hadoop, Spark

예시: 페타바이트급 고객 거래 데이터 분석

C

Canary Deployment

정의: 새로운 모델을 소규모 사용자에게 먼저 배포하여 점진적으로 확대하는 배포 방법

활용: 배포 리스크 최소화

관련 용어: 점진적 배포, Blue-Green 배포, 롤백

예시: 5% → 25% → 50% → 100% 순으로 트래픽을 점진적으로 증가

CI/CD (Continuous Integration/Continuous Deployment)

정의: 코드 통합과 배포를 자동화하는 소프트웨어 개발 방법론

활용: 배포 효율성 향상 및 오류 감소

관련 용어: Jenkins, GitHub Actions, 자동 배포

예시: 코드 커밋 시 자동으로 빌드, 테스트, 배포 수행

Classification (분류)

정의: 입력 데이터를 미리 정의된 카테고리로 분류하는 작업

활용: 이미지 분류, 스팸 필터링, 질병 진단

관련 용어: 이진 분류, 다중 클래스 분류, 레이블

예시: 이메일을 스팸/정상으로 분류

Cloud Computing

정의: 인터넷을 통해 컴퓨팅 리소스를 제공받는 서비스

활용: 인프라 구축 비용 절감 및 확장성 확보

관련 용어: AWS, Azure, GCP, IaaS, PaaS, SaaS

예시: AWS EC2 인스턴스로 모델 학습 서버 구축

Clustering (군집화)

정의: 유사한 데이터를 그룹으로 묶는 비지도 학습 기법

활용: 고객 세분화, 이상 탐지, 데이터 탐색

관련 용어: K-means, DBSCAN, 계층적 군집화

예시: 고객을 구매 패턴에 따라 5 개 그룹으로 분류

CMMI (Capability Maturity Model Integration)

정의: 조직의 프로세스 성숙도를 평가하는 모델

활용: 조직 역량 평가 및 개선

관련 용어: 성숙도 모델, 프로세스 개선

예시: CMMI 5 단계 - 초기, 반복, 정의, 관리, 최적화

CNN (Convolutional Neural Network)

정의: 이미지 처리에 특화된 신경망 구조

활용: 이미지 분류, 객체 탐지, 얼굴 인식

관련 용어: 컨볼루션, 풀링, ResNet, VGG

예시: CNN 을 사용한 고양이/강아지 이미지 분류

Containerization

정의: 애플리케이션과 실행 환경을 컨테이너로 패키징하는 기술

활용: 일관된 실행 환경 및 배포 간소화

관련 용어: Docker, Kubernetes, 컨테이너 오케스트레이션

예시: Docker 로 모델 서빙 애플리케이션 컨테이너화

Cross-validation

정의: 데이터를 여러 폴드로 나누어 모델 성능을 평가하는 기법

활용: 모델의 일반화 성능 평가

관련 용어: K-fold, 검증 세트, 과적합 방지

예시: 5-fold cross-validation 으로 모델 성능 평가

D

Data Drift

정의: 시간이 지남에 따라 입력 데이터의 통계적 특성이 변화하는 현상

활용: 모델 재학습 필요성 판단

관련 용어: 모델 드리프트, 개념 드리프트, 모니터링

예시: 고객 행동 패턴 변화로 입력 데이터 분포 변화

Data Lake

정의: 구조화되지 않은 원시 데이터를 저장하는 대용량 저장소

활용: 다양한 형태의 데이터를 원본 그대로 보존

관련 용어: Data Warehouse, Hadoop, S3

예시: AWS S3 에 로그, 이미지, 센서 데이터를 원본 형식으로 저장

Data Warehouse

정의: 구조화된 데이터를 분석 목적으로 저장하는 저장소

활용: 빠른 쿼리 성능 및 비즈니스 인텔리전스

관련 용어: ETL, OLAP, Data Mart

예시: Snowflake 에서 매출 데이터 분석 쿼리 실행

Deep Learning

정의: 다층 신경망을 사용한 머신러닝 기법

활용: 이미지, 음성, 텍스트 처리

관련 용어: 신경망, CNN, RNN, Transformer

예시: ResNet 으로 이미지 분류 모델 개발

Docker

정의: 컨테이너 기반 가상화 플랫폼

활용: 애플리케이션 패키징 및 배포

관련 용어: 컨테이너, 이미지, Dockerfile

예시: Docker 이미지로 모델 서빙 애플리케이션 배포

DVC (Data Version Control)

정의: 데이터와 모델의 버전을 관리하는 도구

활용: 데이터 버전 관리 및 재현성 확보

관련 용어: 버전 관리, Git, 데이터 리니지

예시: DVC 로 학습 데이터셋의 버전 관리

E

EDM (Evaluate, Direct, Monitor)

정의: IT 거버넌스의 3 대 프로세스 - 평가, 지시, 모니터

활용: IT 거버넌스 프레임워크 구축

관련 용어: ISO 38500, IT 거버넌스, COBIT

예시: EDM 사이클을 통한 AI 프로젝트 거버넌스 관리

Edge Deployment

정의: 모델을 엣지 디바이스에 직접 배포하는 방식

활용: 낮은 지연 시간 및 오프라인 동작

관련 용어: TensorFlow Lite, Edge AI, IoT

예시: 스마트폰에 이미지 분류 모델 배포

Ensemble Learning

정의: 여러 모델을 결합하여 성능을 향상시키는 기법

활용: 예측 정확도 향상

관련 용어: Bagging, Boosting, Stacking

예시: Random Forest, XGBoost, Voting Classifier

ETL (Extract, Transform, Load)

정의: 데이터를 추출, 변환, 적재하는 프로세스

활용: 데이터 통합 및 웨어하우스 구축

관련 용어: 데이터 파이프라인, ELT, 데이터 통합

예시: 여러 소스에서 데이터를 추출하여 변환 후 데이터 웨어하우스에 적재

EU AI Act

정의: 유럽연합의 AI 규제 법안

활용: AI 시스템의 위험 수준별 규제 준수

관련 용어: 고위험 AI, 규제 준수, AI 거버넌스

예시: 고위험 AI 시스템에 대한 엄격한 요구사항 준수

F

F1 Score

정의: 정밀도와 재현율의 조화 평균

활용: 분류 모델 성능 평가

관련 용어: Precision, Recall, 혼동 행렬

예시: $F1\ Score = 2 \times (Precision \times Recall) / (Precision + Recall)$

Fairness

정의: AI 시스템이 모든 그룹에게 공정하게 작동하는 특성

활용: 차별 방지 및 윤리적 AI 구현

관련 용어: Bias, Equity, 공정성 지표

예시: 성별, 인종에 관계없이 동일한 성능을 보이는 AI

Feature Engineering

정의: 원시 데이터로부터 유용한 특성을 생성하는 과정

활용: 모델 성능 향상

관련 용어: Feature Selection, Feature Transformation

예시: 날짜에서 요일, 월, 분기 특성 추출

Feature Store

정의: 머신러닝 특성을 저장하고 관리하는 중앙 저장소

활용: 특성 재사용 및 일관성 보장

관련 용어: Feast, Tecton, Feature Engineering

예시: Feast 로 고객 특성을 저장하고 여러 모델에서 재사용

Fine-tuning

정의: 사전 학습된 모델을 특정 작업에 맞게 추가 학습하는 과정

활용: 소량 데이터로 고성능 달성

관련 용어: Transfer Learning, 사전 학습 모델

예시: BERT 를 고객 리뷰 감정 분석 작업에 맞게 Fine-tuning

G

GCP (Google Cloud Platform)

정의: 구글이 제공하는 클라우드 컴퓨팅 플랫폼

활용: AI/ML 인프라 구축

관련 용어: BigQuery, Vertex AI, Cloud Storage

예시: GCP Vertex AI 로 모델 학습 및 배포

GDPR (General Data Protection Regulation)

정의: 유럽연합의 개인정보 보호 규정

활용: 개인정보 보호 규제 준수

관련 용어: 데이터 프라이버시, 동의 관리, 개인정보 보호

예시: GDPR 준수를 위한 개인정보 처리 동의 절차 구축

Git

정의: 분산 버전 관리 시스템

활용: 코드 버전 관리 및 협업

관련 용어: GitHub, GitLab, 버전 관리

예시: Git 으로 모델 코드 버전 관리

GPU (Graphics Processing Unit)

정의: 병렬 처리에 특화된 프로세서

활용: 딥러닝 모델 학습 가속

관련 용어: CUDA, 분산 학습, GPU 클러스터

예시: NVIDIA A100 GPU 로 대규모 모델 학습

GPT (Generative Pre-trained Transformer)

정의: OpenAI 가 개발한 생성형 대규모 언어 모델

활용: 텍스트 생성, 질의응답, 대화

관련 용어: LLM, Transformer, 생성형 AI

예시: GPT-4 를 사용한 고객 응대 챗봇 구축

Grafana

정의: 모니터링 데이터 시각화 플랫폼

활용: 실시간 모니터링 대시보드 구축

관련 용어: Prometheus, 시각화, 대시보드

예시: Grafana 로 모델 성능 메트릭 실시간 모니터링

gRPC

정의: 고성능 RPC(원격 프로시저 호출) 프레임워크

활용: 마이크로서비스 간 통신

관련 용어: REST API, Protocol Buffers, API

예시: gRPC 로 모델 서빙 API 구현

H

Hadoop

정의: 대용량 데이터 분산 처리 프레임워크

활용: 빅데이터 배치 처리

관련 용어: MapReduce, HDFS, Spark

예시: Hadoop 으로 페타바이트급 로그 데이터 처리

Hyperparameter Tuning

정의: 모델의 최적 하이퍼파라미터를 찾는 과정

활용: 모델 성능 최적화

관련 용어: Grid Search, Random Search, Optuna

예시: Optuna 로 학습률, 배치 크기 등의 최적 값 탐색

I

Inference

정의: 학습된 모델로 새 데이터에 대한 예측 수행

활용: 실제 서비스에서 모델 사용

관련 용어: Prediction, Model Serving, 추론 시간

예시: 실시간 추천 서비스에서 모델 추론 수행

IoT (Internet of Things)

정의: 사물들이 인터넷으로 연결되어 데이터를 주고받는 기술

활용: 센서 데이터 수집 및 스마트 디바이스 제어

관련 용어: 센서, Edge Computing, 실시간 데이터

예시: 공장 설비의 IoT 센서 데이터로 예측 정비 수행

ISO 27001

정의: 정보 보안 관리 시스템 국제 표준

활용: 정보 보안 체계 구축 및 인증

관련 용어: 보안, 인증, ISMS

예시: ISO 27001 인증을 통한 데이터 보안 관리 체계 구축

ISO 38500

정의: IT 거버넌스 국제 표준

활용: IT 거버넌스 프레임워크 구축

관련 용어: EDM, 거버넌스, IT 관리

예시: ISO 38500 기반 AI 거버넌스 체계 수립

ISO 42001

정의: AI 관리 시스템 국제 표준

활용: AI 거버넌스 및 관리 체계 구축

관련 용어: AI 거버넌스, AI 관리, 인증

예시: ISO 42001 인증을 통한 AI 관리 시스템 구축

J

Jenkins

정의: CI/CD 자동화 도구

활용: 빌드, 테스트, 배포 자동화

관련 용어: CI/CD, 자동화, 파이프라인

예시: Jenkins 로 모델 재학습 및 배포 파이프라인 자동화

Jupyter Notebook

정의: 대화형 컴퓨팅 환경을 제공하는 웹 애플리케이션

활용: 데이터 분석, 모델 개발, 문서화

관련 용어: Python, 데이터 과학, 프로토타이핑

예시: Jupyter Notebook 으로 탐색적 데이터 분석 수행

K

Kafka

정의: 분산 스트리밍 플랫폼

활용: 실시간 데이터 파이프라인 구축

관련 용어: 메시지 큐, 스트리밍, 이벤트 처리

예시: Kafka 로 실시간 거래 데이터 스트리밍 처리

Keras

정의: 고수준 딥러닝 API

활용: 신경망 모델 빠른 프로토타이핑

관련 용어: TensorFlow, 딥러닝, API

예시: Keras 로 CNN 이미지 분류 모델 구현

KPI (Key Performance Indicator)

정의: 핵심 성과 지표

활용: 목표 달성도 측정 및 평가

관련 용어: 성과 관리, 메트릭, 목표

예시: AI 프로젝트의 ROI, 정확도, 응답 시간을 KPI 로 설정

Kubeflow

정의: Kubernetes 기반 MLOps 플랫폼

활용: ML 파이프라인 구축 및 관리

관련 용어: Kubernetes, MLOps, 파이프라인

예시: Kubeflow 로 모델 학습부터 배포까지 파이프라인 구축

Kubernetes

정의: 컨테이너 오케스트레이션 플랫폼

활용: 컨테이너 배포, 확장, 관리 자동화

관련 용어: Docker, 컨테이너, 오케스트레이션

예시: Kubernetes 로 모델 서빙 마이크로서비스 관리

L

LIME (Local Interpretable Model-agnostic Explanations)

정의: 개별 예측에 대한 설명을 제공하는 XAI 기법

활용: 모델 의사결정 설명

관련 용어: XAI, SHAP, 설명 가능성

예시: LIME 으로 대출 거절 이유 설명

LLM (Large Language Model)

정의: 대규모 텍스트 데이터로 학습된 언어 모델

활용: 텍스트 생성, 질의응답, 번역

관련 용어: GPT, BERT, Transformer

예시: GPT-4, Claude, LLaMA 등의 대규모 언어 모델

LSTM (Long Short-Term Memory)

정의: 장기 의존성을 학습할 수 있는 순환 신경망

활용: 시계열 예측, 자연어 처리

관련 용어: RNN, 시계열, 순차 데이터

예시: LSTM 으로 주가 예측 모델 구축

M

Machine Learning

정의: 데이터로부터 패턴을 학습하는 AI 기술

활용: 예측, 분류, 군집화 등

관련 용어: 지도 학습, 비지도 학습, 강화 학습

예시: 고객 이탈 예측, 상품 추천, 이미지 분류

MLflow

정의: 머신러닝 라이프사이클 관리 플랫폼

활용: 실험 추적, 모델 관리, 배포

관련 용어: MLOps, Model Registry, 실험 관리

예시: MLflow 로 하이퍼파라미터, 메트릭, 모델 버전 관리

MLOps (Machine Learning Operations)

정의: 머신러닝 모델의 개발, 배포, 운영을 자동화하는 방법론

활용: 모델 라이프사이클 관리

관련 용어: DevOps, CI/CD, 모델 모니터링

예시: MLOps 파이프라인으로 모델 재학습 및 배포 자동화

Model Registry

정의: 학습된 모델을 중앙에서 관리하는 시스템

활용: 모델 버전 관리 및 추적

관련 용어: MLflow, 버전 관리, 모델 스테이징

예시: MLflow Model Registry 에 프로덕션, 스테이징, 아카이브 모델 관리

Model Serving

정의: 학습된 모델을 서비스로 제공하는 것

활용: 실시간 예측 서비스 제공

관련 용어: Inference, API, TensorFlow Serving

예시: TensorFlow Serving 으로 REST API 형태의 예측 서비스 제공

Monitoring

정의: 시스템의 상태와 성능을 지속적으로 관찰

활용: 문제 조기 발견 및 대응

관련 용어: Prometheus, Grafana, 로깅

예시: 모델 성능, 데이터 드리프트, 시스템 리소스 모니터링

N

Neural Network

정의: 인간의 뇌 구조를 모방한 계산 모델

활용: 복잡한 패턴 학습

관련 용어: 딥러닝, 퍼셉트론, 활성화 함수

예시: 다층 신경망으로 이미지 인식 모델 구축

NLP (Natural Language Processing)

정의: 자연어를 처리하고 이해하는 기술

활용: 텍스트 분석, 챗봇, 번역

관련 용어: BERT, GPT, 토큰화, 임베딩

예시: 고객 문의 자동 분류, 감정 분석, 기계 번역

NumPy

정의: 파이썬 수치 연산 라이브러리

활용: 행렬 연산, 수학적 계산

관련 용어: Python, Pandas, 배열

예시: NumPy로 행렬 연산 및 데이터 전처리

O

Object Detection

정의: 이미지에서 객체의 위치와 종류를 식별하는 기술

활용: 자율주행, 보안, 품질 검사

관련 용어: YOLO, R-CNN, Bounding Box

예시: YOLO로 도로 위 차량, 보행자 탐지

OLAP (Online Analytical Processing)

정의: 다차원 데이터 분석을 위한 기술

활용: 비즈니스 인텔리전스, 의사결정 지원

관련 용어: Data Warehouse, Cube, 다차원 분석

예시: 연도별, 지역별, 제품별 매출 다차원 분석

Ollama

정의: 로컬에서 LLM을 실행하는 오픈소스 프레임워크

활용: 온프레미스 LLM 구축 및 데이터 보안

관련 용어: Local LLM, LLaMA, 데이터 보안

예시: Ollama로 기업 내부 데이터를 사용하는 챗봇 구축

Optuna

정의: 하이퍼파라미터 최적화 프레임워크

활용: 자동 하이퍼파라미터 튜닝

관련 용어: Hyperparameter Tuning, AutoML

예시: Optuna 로 최적의 학습률, 배치 크기 자동 탐색

Overfitting

정의: 학습 데이터에 과적합되어 일반화 성능이 떨어지는 현상

활용: 모델 성능 개선

관련 용어: 정규화, 드롭아웃, 조기 종료

예시: 검증 데이터에서 성능이 계속 떨어질 때 학습 중단

P

Pandas

정의: 파이썬 데이터 분석 라이브러리

활용: 데이터 조작 및 분석

관련 용어: DataFrame, Python, NumPy

예시: Pandas DataFrame 으로 CSV 데이터 로드 및 전처리

Pipeline

정의: 데이터 처리나 ML 워크플로우를 자동화한 것

활용: 반복 작업 자동화

관련 용어: MLOps, CI/CD, 워크플로우

예시: 데이터 수집→전처리→학습→배포 자동화 파이프라인

Precision

정의: 양성으로 예측한 것 중 실제 양성인 비율

활용: 분류 모델 성능 평가

관련 용어: Recall, F1 Score, 혼동 행렬

예시: $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$

Prometheus

정의: 오픈소스 모니터링 및 알림 시스템

활용: 시스템 및 애플리케이션 모니터링

관련 용어: Grafana, 메트릭, 시계열 데이터

예시: Prometheus 로 모델 성능 메트릭 수집

PyTorch

정의: 페이스북이 개발한 딥러닝 프레임워크

활용: 신경망 모델 개발 및 연구

관련 용어: TensorFlow, 딥러닝, GPU

예시: PyTorch 로 Transformer 모델 구현

Python

정의: AI/ML 에 널리 사용되는 프로그래밍 언어

활용: 데이터 분석, 모델 개발, 자동화

관련 용어: NumPy, Pandas, Scikit-learn

예시: Python 으로 데이터 전처리 및 모델 학습 코드 작성

Q

Quality Assurance (QA)

정의: 제품이나 서비스의 품질을 보장하는 활동

활용: 모델 및 시스템의 품질 검증

관련 용어: 테스트, 검증, 품질 관리

예시: 모델 배포 전 성능, 편향성, 보안 테스트 수행

R

RACI Matrix

정의: 책임 할당 매트릭스 (Responsible, Accountable, Consulted, Informed)

활용: 역할과 책임의 명확화

관련 용어: 책임 관리, 프로젝트 관리

예시: AI 프로젝트의 각 작업별 RACI 정의

Random Forest

정의: 다수의 결정 트리를 앙상블한 모델

활용: 분류 및 회귀 문제 해결

관련 용어: Decision Tree, Ensemble, Bagging

예시: Random Forest 로 고객 이탈 예측

Recall (재현율)

정의: 실제 양성 중 모델이 양성으로 예측한 비율

활용: 분류 모델 성능 평가

관련 용어: Precision, F1 Score, Sensitivity

예시: $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$

Regression (회귀)

정의: 연속적인 수치 값을 예측하는 작업

활용: 가격 예측, 수요 예측, 매출 예측

관련 용어: 선형 회귀, 로지스틱 회귀

예시: 주택 가격 예측, 매출 예측

Reinforcement Learning (강화 학습)

정의: 보상을 최대화하는 행동을 학습하는 기법

활용: 게임 AI, 로봇 제어, 최적화

관련 용어: Q-learning, DQN, 정책 최적화

예시: AlphaGo, 자율주행 차량 제어

REST API

정의: HTTP 기반의 웹 서비스 API

활용: 시스템 간 통신 및 서비스 제공

관련 용어: HTTP, JSON, gRPC

예시: REST API로 모델 예측 서비스 제공

ResNet

정의: 잔차 연결을 사용하는 CNN 아키텍처

활용: 이미지 분류, 객체 탐지

관련 용어: CNN, Skip Connection, 딥러닝

예시: ResNet-50으로 이미지 분류 모델 개발

RNN (Recurrent Neural Network)

정의: 순차 데이터 처리에 특화된 신경망

활용: 시계열 예측, 자연어 처리

관련 용어: LSTM, GRU, 순차 데이터

예시: RNN으로 주가 예측, 텍스트 생성

ROI (Return on Investment)

정의: 투자 대비 수익률

활용: AI 프로젝트의 경제적 가치 평가

관련 용어: 비용 편익 분석, 투자 효과

예시: AI 도입으로 연간 10 억원 절감, 투자 5 억원 → ROI 200%

Rollback

정의: 시스템을 이전 상태로 되돌리는 작업

활용: 배포 실패 시 빠른 복구

관련 용어: 배포, 버전 관리, Blue-Green Deployment

예시: 새 모델 배포 후 오류 발생 시 이전 버전으로 자동 복구

S

SageMaker

정의: AWS 의 완전 관리형 ML 플랫폼

활용: 모델 개발, 학습, 배포

관련 용어: AWS, MLOps, 클라우드

예시: SageMaker 로 모델 학습 및 엔드포인트 배포

Scikit-learn

정의: 파이썬 머신러닝 라이브러리

활용: 머신러닝 모델 개발

관련 용어: Python, 머신러닝, 알고리즘

예시: Scikit-learn 으로 랜덤 포레스트, SVM 모델 구현

SHAP (SHapley Additive exPlanations)

정의: 각 특성의 예측 기여도를 계산하는 XAI 기법

활용: 모델 해석 및 설명

관련 용어: XAI, LIME, 설명 가능성

예시: SHAP 으로 각 특성이 예측에 미친 영향 시각화

Spark

정의: 대규모 데이터 분산 처리 프레임워크

활용: 빅데이터 처리 및 분석

관련 용어: Hadoop, PySpark, 분산 처리

예시: Spark 로 테라바이트급 데이터 병렬 처리

SQL (Structured Query Language)

정의: 데이터베이스 질의 언어

활용: 데이터 조회, 삽입, 수정, 삭제

관련 용어: Database, 관계형 데이터베이스

예시: SQL 로 고객 데이터 조회 및 집계

Supervised Learning (지도 학습)

정의: 레이블이 있는 데이터로 모델을 학습하는 방식

활용: 분류, 회귀 문제 해결

관련 용어: Classification, Regression, 레이블

예시: 레이블된 이미지 데이터로 분류 모델 학습

T

TensorFlow

정의: 구글이 개발한 딥러닝 프레임워크

활용: 신경망 모델 개발 및 배포

관련 용어: Keras, PyTorch, 딥러닝

예시: TensorFlow 로 CNN 모델 구현 및 학습

TensorFlow Serving

정의: TensorFlow 모델 서빙 시스템

활용: 프로덕션 환경에서 모델 배포

관련 용어: Model Serving, Inference, REST API

예시: TensorFlow Serving 으로 모델을 REST API 로 제공

Test Data

정의: 모델의 최종 성능 평가용 데이터

활용: 모델의 일반화 성능 측정

관련 용어: Train Data, Validation Data

예시: 전체 데이터의 20%를 테스트 데이터로 분리

Training

정의: 데이터로 모델의 파라미터를 최적화하는 과정

활용: AI 모델 개발

관련 용어: 학습, 최적화, 역전파

예시: 10,000 개 이미지로 100 에포크 동안 모델 학습

Transfer Learning

정의: 한 작업에서 학습한 지식을 다른 작업에 적용

활용: 학습 시간 단축 및 성능 향상

관련 용어: Fine-tuning, 사전 학습 모델

예시: ImageNet 사전 학습 모델을 의료 이미지에 적용

Transformer

정의: 어텐션 메커니즘 기반의 신경망 아키텍처

활용: 자연어 처리, 이미지 처리

관련 용어: BERT, GPT, Attention

예시: Transformer 로 기계 번역 모델 구현

Transparency

정의: AI 시스템의 작동 방식을 명확히 이해할 수 있는 특성

활용: 신뢰성 확보 및 책임성 보장

관련 용어: Explainability, 설명 가능성

예시: 모델 카드로 모델의 작동 방식 문서화

U

Underfitting

정의: 모델이 데이터의 패턴을 충분히 학습하지 못한 상태

활용: 모델 복잡도 조정

관련 용어: Overfitting, 모델 복잡도, 일반화

예시: 너무 단순한 모델로 복잡한 패턴 학습 실패

Unsupervised Learning (비지도 학습)

정의: 레이블 없는 데이터로 패턴을 학습하는 방식

활용: 군집화, 차원 축소, 이상 탐지

관련 용어: Clustering, K-means, PCA

예시: 고객 구매 패턴 기반 세분화 (군집화)

V

Validation Data

정의: 모델의 하이퍼파라미터 튜닝 및 중간 평가용 데이터

활용: 과적합 방지 및 모델 선택

관련 용어: Train Data, Test Data, Cross-validation

예시: 전체 데이터의 20%를 검증 데이터로 분리

Version Control

정의: 코드, 데이터, 모델의 변경 이력을 관리하는 체계

활용: 협업 및 변경 추적

관련 용어: Git, DVC, 버전 관리

예시: Git 으로 코드 버전 관리, DVC 로 데이터 버전 관리

W

Webhook

정의: 이벤트 발생 시 자동으로 HTTP 요청을 보내는 메커니즘

활용: 시스템 간 실시간 통합

관련 용어: API, 이벤트 기반, 자동화

예시: 모델 배포 완료 시 Slack 으로 알림 전송

X

XAI (eXplainable AI)

정의: AI 의 의사결정 과정을 설명 가능하게 만드는 기술

활용: AI 투명성 및 신뢰성 확보

관련 용어: SHAP, LIME, 설명 가능성

예시: SHAP 으로 대출 승인/거절 이유 설명

XGBoost

정의: 그래디언트 부스팅 기반의 앙상블 모델

활용: 분류 및 회귀 문제 해결

관련 용어: Gradient Boosting, Ensemble, 트리 모델

예시: XGBoost 로 고객 이탈 예측 모델 구축

Y

YOLO (You Only Look Once)

정의: 실시간 객체 탐지 알고리즘

활용: 빠른 객체 탐지

관련 용어: Object Detection, CNN, 실시간 처리

예시: YOLOv5 로 실시간 보행자 탐지

Z

Zero-shot Learning

정의: 학습하지 않은 클래스를 예측할 수 있는 학습 방식

활용: 소량 데이터 환경에서 모델 활용

관련 용어: Few-shot Learning, Transfer Learning

예시: 고양이만 학습한 모델이 호랑이도 인식

부록

용어 통계

- 한글 용어: 61 개
- 영문 용어: 117 개
- 총 용어 수: 178 개

카테고리별 분류

1. MLOps 및 인프라 (35 개): CI/CD, Docker, Kubernetes, MLflow, Pipeline, Model Registry, Model Serving, Auto-scaling 등
2. AI/ML 기술 (45 개): Machine Learning, Deep Learning, CNN, RNN, Transformer, BERT, GPT, NLP, Computer Vision 등
3. 데이터 관리 (28 개): Data Lake, Data Warehouse, Data Pipeline, ETL, Data Drift, Data Quality, Feature Store 등
4. 거버넌스 및 윤리 (22 개): Governance, Risk Management, Fairness, Bias, XAI, SHAP, LIME, Transparency 등
5. 표준 및 규제 (8 개): ISO 42001, ISO 38500, ISO 27001, GDPR, EU AI Act, CMMI 등
6. 도구 및 플랫폼 (25 개): TensorFlow, PyTorch, Scikit-learn, AWS, GCP, SageMaker, Prometheus, Grafana 등
7. 아키텍처 및 배포 (15 개): Microservices, API, REST API, gRPC, Canary Deployment, Rollback, Edge Deployment 등

참고 문헌

- ISO/IEC 42001:2023 - AI Management System
- ISO/IEC 38500:2015 - IT Governance
- ISO/IEC 27001:2013 - Information Security Management
- EU AI Act (2024)
- GDPR (General Data Protection Regulation)
- AiNex 프로젝트 컨설팅 방법론 문서

문서 끝

> 본 용어집은 AiNex AI 컨설팅 프로젝트의 공식 용어 가이드로, 프로젝트 수행 시 표준 용어로 활용되어야 합니다.