

# 설명가능한 인공지능(XAI) 기술 동향과 데이터 산업의 시장 전망

## 목 차

---

### 제1장 설명 가능한 인공지능(XAI) 개요 및 기술 동향

1. 설명 가능한 인공지능(XAI) 기술 개요
  - 1-1. XAI(eXplainable AI) 등장 배경과 개념
    - 1-1-1. XAI 등장 배경
    - 1-1-2. AI의 급속한 확산에 따른 부작용
      - (1) 인공지능의 불쾌한 골짜기(Uncanny Valley)
        - 가. 딥러닝의 진화
        - 나. 현재 인공지능이 지닌 문제점, 블랙박스의 미스터리
        - 다. 인공지능의 대표적 오류 사례
      - (2) 인공지능의 편향성(Bias) 문제
        - 가. AI의 의사결정 지원과 편향성(Bias)
        - 나. 데이터의 편향성(Bias)
        - 다. 데이터 편향 유형
        - 라. 인공지능의 편향성(Bias) 문제 해결 방안
      - (3) 인공지능의 편향성(Bias) 문제 해결 방안
        - 가. 데이터 경제(Data Economy) 시대
        - 나. 데이터 소유와 독점
    - 1-1-3. 미래의 인공지능 알고리즘
      - (1) AI 부작용이나 위험성에 대한 해결 방안
      - (2) 미래 인공지능 알고리즘
    - 1-1-4. XAI 개요
      - (1) XAI의 개념
      - (2) XAI의 필요성
  - 1-2. AI 2.0 시대 XAI 기술 개요
    - 1-2-1. AI 2.0 시대
    - 1-2-2. 설명 가능한 인공지능의 작용 방식
      - (1) 기존 학습 모델 변형: 심층신경망에 설명 가능성 부여하기 다윈AI 생성 합성(Generative Synthesis) 기술
      - (2) DAPRA XAI 전략, 기본 설계부터 인간이 이해할 수 있는 구조로 신경망을 만드는 방식
        - 가. 심층설명학습(deep explanation)
        - 나. 해석 가능한 모델(interpretable models)
        - 다. 모델 귀납(model induction)
      - (3) 학습모델간 비교
2. XAI 기술 동향 및 개발 현황
  - 2-1. XAI 프로세스 개요
    - 2-1-1. XAI 프로세스
    - 2-1-2. XAI의 접근 방법
  - 2-2. XAI 개발을 위한 기술적 접근
    - 2-2-1. 신경회로망 노드에 설명라벨 붙이기
    - 2-2-2. 의사결정트리를 이용한 설명모델 만들기
    - 2-2-3. 통계적 방법을 이용하여 설명모델 유추하기

## 2-3. XAI 기대효과 및 시사점

### 2-3-1. XAI 기대효과

- (1) 경계 사례와 데이터 편향성을 탐지·제거함으로써 성능 향상
- (2) 모델 정확성 및 성능 개선
- (3) 신뢰성 확보

### 2-3-2. 시사점

## 3. 설명가능한 인공지능 알고리즘 및 XAI 개발 동향

### 3-1. 설명가능한 인공지능 알고리즘

- 3-1-1. 부분 의존 구성(Partial Dependence Plots, PDP)
- 3-1-2. 개별 조건 예측(Individual Conditional Expectations, ICE)
- 3-1-3. 민감도 분석(Sensitivity Analysis, SA)
- 3-1-4. 계층별 관련도 전파법(Layer-wise Relevance Propagation, LRP)
- 3-1-5. 일부 해석 모델(Local Interpretable Model-agnostic Explanation, LIME)
- 3-1-6. 첨가 요인 민감도(Sharply Additive Explanations, SHAP)

### 3-2. 활용 분야

- 3-2-1. 금융 / 핀테크 분야 서비스
- 3-2-2. 의료 / 헬스케어 분야 서비스
- 3-2-3. 자율주행 자동차
- 3-2-4. 제조

### 3-3. XAI 산업 동향 및 기술 개발 현황

#### 3-3-1. XAI 기술 개발

- (1) 미국 국방성 산하 방위고등연구계획국(DARPA)
- (2) IBM
- (3) 구글
- (4) 페이스북(Facebook)
- (5) 심머신(simMachines, Inc)
- (6) 국내 XAI 연구

#### 3-3-2. XAI 기술의 특허 동향

## 4. 주요 AI 알고리즘 트렌드

### 4-1. 제로샷 학습(zero-shot learning)

- 4-1-1. 제로샷 학습(zero-shot learning) 개념
- 4-1-2. 제로샷 학습(zero-shot learning) 원리
- 4-1-3. 제로샷 방법론

### 4-2. 생성적 적대 신경망(Generative Adversarial Network)

#### 4-2-1. GAN(Generative Adversarial Network) 개요 및 학습 방법

- (1) GAN(Generative Adversarial Network, 적대적 생성 신경망) 개요 및 정의  
가. GAN 개요  
나. GAN 구조
  - ① 학습데이터
  - ② 생성자(generator) 네트워크
  - ③ 판별자(discriminator) 네트워크

#### (2) 적대적 학습방법

#### 4-2-2. GAN 응용 모델과 적용 사례

- (1) CGAN(Conditional GAN)
- (2) InfoGAN
- (3) Laplacian GAN
- (4) DCGAN(Deep Convolutional Generative Adversarial Networks)
- (5) DiscoGAN

### 4-3. 강화학습(Reinforcement Learning)

#### 4-3-1. 강화학습(Reinforcement Learning) 개요

- (1) 강화학습(Reinforcement Learning)의 개요  
가. MDP(Markov Decision Process) 방식

- 나. DQN(Deep Q-Network)
- 4-3-2. 강화학습(Reinforcement Learning)의 특징
- 4-4. 전이학습(transfer learning)
- 4-4-1. 전이학습(transfer learning) 개요
  - (1) 전이학습(Transfer learning) 개념
  - (2) 전이학습 특징
- 4-4-2. 전이학습 알고리즘

## 제2장 데이터 경제 시대 미래 비즈니스 생태계를 위한 데이터 활용

1. 데이터 경제 시대 미래 비즈니스 생태계
  - 1-1. 비대면 시대
    - 1-1-1. 포스트 코로나 시대 디지털 전환
    - 1-1-2. 비대면 시대, 인공지능(AI)과 데이터 아키텍처의 미래
      - (1) 인공지능(AI)과 비대면
      - (2) 비대면 시대, 인공지능(AI)과 데이터 아키텍처
  - 1-2. 데이터 경제(data economics) 시대 데이터 역할
    - 1-2-1. 데이터 경제 시대의 개요
    - 1-2-2. 데이터 오너십(data ownership)
      - (1) 데이터 오너십(data ownership) 개요
      - (2) 데이터 소유권 문제
        - 가. 데이터 소유권 개념
        - 나. 데이터 소유권에 대한 기준
        - 다. 데이터 거래
2. 인공지능 시대 데이터 활용
  - 2-1. 데이터 산업
    - 2-1-1. 데이터옵스(DataOps)
      - (1) 데이터옵스(DataOps) 개념
      - (2) 데이터옵스(DataOps)의 아키텍처
      - (3) 데이터옵스의 운영 프로세스
    - 2-1-2. AI옵스(AIOps)
      - (1) AIOps 개념
      - (2) AI옵스 활용
      - (3) AI옵스 시장 전망
  - 2-2. 글로벌 데이터 시장과 각국의 정책 현황
    - 2-2-1. 데이터 시장
    - 2-2-2. 데이터 경제 정책 현황
      - (1) 미국
      - (2) 유럽연합(EU)
      - (3) 중국
      - (4) 일본
      - (5) 우리나라

참고문헌

### 그림목차

- [그림 1] XAI의 필요성
- [그림 2] 불쾌한 골짜기(Uncanny Valley)
- [그림 3] 다양한 스케일링 방법 비교
- [그림 4] 신경망 기본 모델(a)과 다중 목표 최적화를 위한 진화 알고리즘 프레임워크(b)
- [그림 5] 심층신경망의 구조와 훈련
- [그림 6] 블랙박스 문제 분류
- [그림 7] A Brief History of Machine Learning Models Explainability(성능 vs 설명)
- [그림 8] 기계학습의 오류 원인

- [그림 9] 분산형 AI 플랫폼 비전
- [그림 10] AI 편향을 줄이기 위한 엔지니어링 원칙
- [그림 11] 미래 기술의 시너지
- [그림 12] 글로벌 산업별 데이터 활용
- [그림 13] 글로벌 데이터센터 시장 규모
- [그림 14] 데이터 경제의 가치 창출 체계
- [그림 15] AI 기반 의사결정(Decision Making)
- [그림 16] 현재의 인공지능과 XAI
- [그림 17] eXplainable AI시스템의 표현
- [그림 18] XAI 개발 과제
- [그림 19] 블랙박스(Black box)로 인해 설명력이 낮아진 인공지능
- [그림 20] XAI의 모델 해석 성과
- [그림 21] XAI 개발을 위한 기술적 접근
- [그림 22] XAI 프레임 워크
- [그림 23] 역합성곱 신경망 구조 예시
- [그림 24] 반복적인 모델 설명
- [그림 25] AND-OR 그래프를 이용한 이미지 분류
- [그림 26] AI의 블랙박스(Black-Box)와 설명가능한 AI
- [그림 27] XAI 프로그램의 구조
- [그림 28] XAI의 접근 방법
- [그림 29] 인공신경망의 설명가능한 노드에 대한 레이블 예시
- [그림 30] XAI의 모니터링과 분석 과정
- [그림 31] 정확성과 설명력의 트레이드오프(Trade off)
- [그림 32] 설명가능한 AI 모델의 분류
- [그림 33] XAI 기술 및 전략
- [그림 34] AI 설명 가능성의 세 단계
- [그림 35] LIME 이미지 분류
- [그림 36] XAI 프레임 워크
- [그림 37] 딥러닝의 사물인식 과정에 XAI가 적용될 경우
- [그림 38] XAI에 대한 개념과 접근방식
- [그림 39] 머신러닝 예측 모델에 설명 가능성 부여
- [그림 40] XAI 과제
- [그림 41] 딥러닝의 복잡성
- [그림 42] Industrie 4.0을 위한 XAI
- [그림 43] 설명가능한 AI 설명- 2단계 접근 방식
- [그림 44] DARPA의 설명가능 인공지능 개발 방향
- [그림 45] 기업들이 AI 도입을 망설이는 이유(중복 응답 가능)
- [그림 46] 더 효율적인 CNN, EfficientNet
- [그림 47] 설명가능한 딥러닝 프레임워크
- [그림 48] Facebook의 기계학습 시스템
- [그림 49] Cognilytica AI Positioning Matrix+ ㄷTM
- [그림 50] 딥러닝을 활용한 물체 감지 구조
- [그림 51] 설명가능한 AI기술의 분야·국가별 특허 동향
- [그림 52] 제로샷 학습(zero-shot learning)
- [그림 53] 전이학습(Transfer learning) vs 제로샷 학습(zero-shot learning)
- [그림 54] 구글 신경망 기계번역 시스템의 구조
- [그림 55] 임베딩 기반 방법을 사용한 제로샷 학습
- [그림 56] 생성 모델 기반 방법을 사용한 제로샷 학습
- [그림 57] GAN을 사용하여 속성 벡터에서 이미지 특징 얻기
- [그림 58] generative model의 분류
- [그림 59] Fake and real images
- [그림 60] GAN의 개념도
- [그림 61] GAN의 학습 방법
- [그림 62] Generative Adversarial Network

- [그림 63] Generative model
- [그림 64] 판별자(discriminator) 네트워크
- [그림 65] Adversarial Nets Framework
- [그림 66] Generative Network
- [그림 67] Discriminator Network
- [그림 68] CGAN의 얼굴인식 과정
- [그림 69] CGAN(Conditional GAN)
- [그림 70] InfoGAN 및 Vanilla GAN의 아키텍처
- [그림 71] InfoGAN Implementation
- [그림 72] Laplacian GAN
- [그림 73] DCGAN Architecture
- [그림 74] 기존 GAN Architecture
- [그림 75] DCGAN
- [그림 76] 선택기 신경망과 생성기 신경망
- [그림 77] DiscoGAN 사용 예시
- [그림 78] 강화학습(Reinforcement learning)
- [그림 79] 환경과 상호작용을 통한 강화학습 구조
- [그림 80] 강화학습 프레임워크(Reinforcement Learning Framework)
- [그림 81] 로봇에 적용된 DQN
- [그림 82] q-learning
- [그림 83] 마르코프 결정과정 문제(Markov Decision Process, MDP)
- [그림 84] 딥마인드 DQN 구조
- [그림 85] 미분 가능 신경컴퓨터의 아키텍처 구조
- [그림 86] 강화와 처벌
- [그림 87] 강화학습 시스템 구조
- [그림 88] 전이학습(transfer learning)
- [그림 89] 전통 기계학습과 전이학습의 비교
- [그림 90] 패턴인식(pattern recognition) 프로세스
- [그림 91] PathNet과 Stepwise Pathnet의 비교
- [그림 92] ICT impact realtionships
- [그림 93] 2019년 디지털 혁신 트렌드
- [그림 94] 멀티 클라우드 아키텍처
- [그림 95] 세계 인공지능 헬스케어 시장 규모 2016-2023
- [그림 96] 인공지능의 핵심 영역
- [그림 97] 인공지능(AI) 및 빅데이터
- [그림 98] 데이터 경제의 가치창출 체계
- [그림 99] 데이터 경제(Data Economy) Framework
- [그림 100] 하루동안 생산되는 데이터 양
- [그림 101] 블록체인과 데이터 경제
- [그림 102] 데이터 경제 시스템
- [그림 103] 마이데이터(My Data)의 소유자
- [그림 104] 데이터 소유자
- [그림 105] 데이터 소유권, 보안, 애플리케이션의 관계
- [그림 106] 데이터 값 주기
- [그림 107] 개인 데이터 생태계
- [그림 108] 데이터 소유권 및 관리
- [그림 109] 데이터 거래 절차
- [그림 110] 데이터옵스(DataOps) 개요
- [그림 111] 데이터옵스(DataOps) 아키텍처
- [그림 112] 데이터옵스(DataOps) 라이프사이클
- [그림 113] 데이터 옵스를 사용한 통합 접근 방식
- [그림 114] 머신러닝과 DataOps 사례
- [그림 115] AIOps 접근 방식
- [그림 116] IT 운영 관리에 통찰력을 제공하는 AIOps 플랫폼

- [그림 117] AI오픈스 플랫폼 시각화
- [그림 118] AI오픈스 플랫폼의 논리적 구조
- [그림 119] 국내 빅데이터 및 분석 시장 전망
- [그림 120] 데이터 수집 체계
- [그림 121] 영국의 데이터 포털 사이트
- [그림 122] 데이터 활용을 둘러싼 일본 정책 추진 현황
- [그림 123] 한국판 뉴딜의 구조와 추진체계

**표목차**

- [표 1] 인공지능 개발의 진화와 설명 가능한 AI 워크플로우
- [표 2] 다양한 편향 요인이 반영되는 인공지능과 현재 머신러닝의 워크플로우 구조
- [표 3] 기계학습 시스템 및 XAI 개요
- [표 4] 인공지능의 편향 사례 및 데이터 편향(Bias)
- [표 5] 편향(Bias)의 5가지 종류 및 AI시스템의 편향성 발견을 위한 프로세스
- [표 6] 데이터 경제를 주도하는 GAFAs
- [표 7] 다양한 분야에서의 인공지능 오류
- [표 8] DARPA의 AI Next Campaign 연구 주제
- [표 9] XAI 연구 방향 및 신뢰성 있는 인공지능을 위한 프레임워크
- [표 10] XAI 개요 및 운용 체계
- [표 11] XAI 개발을 위한 기술적 접근 및 주요 과제
- [표 12] XAI 기반 기술 분류와 설명가능한 인공지능(XAI) 예시
- [표 13] DARPA의 XAI 효율적 설명의 평가지표
- [표 14] 알파고(AlphaGo)의 진화
- [표 15] 설명가능한 인공지능 적용 사례 및 모델 성과와 해석 가능성
- [표 16] 의료분야에서 XAI 적용
- [표 17] AI 오픈스케일의 흐름 및 강점
- [표 18] AI기술의 취약점
- [표 19] 머신러닝의 학습 방법
- [표 20] 전이학습(Transfer learning) 워크플로우 및 응용 분야
- [표 21] 의료 빅데이터 활용 강화 분야
- [표 22] 데이터 가치사슬
- [표 23] 데이터 가치 창출 및 데이터 경제 활성화 기대 효과
- [표 24] 해외 데이터 경제 동향
- [표 25] 빅데이터 관련 중국 정부 정책 및 주요 내용 정리