원자력분야 인공지능 기술 활용을 위한 규제 고려사항

<< 한국원자력통제기술원 >>

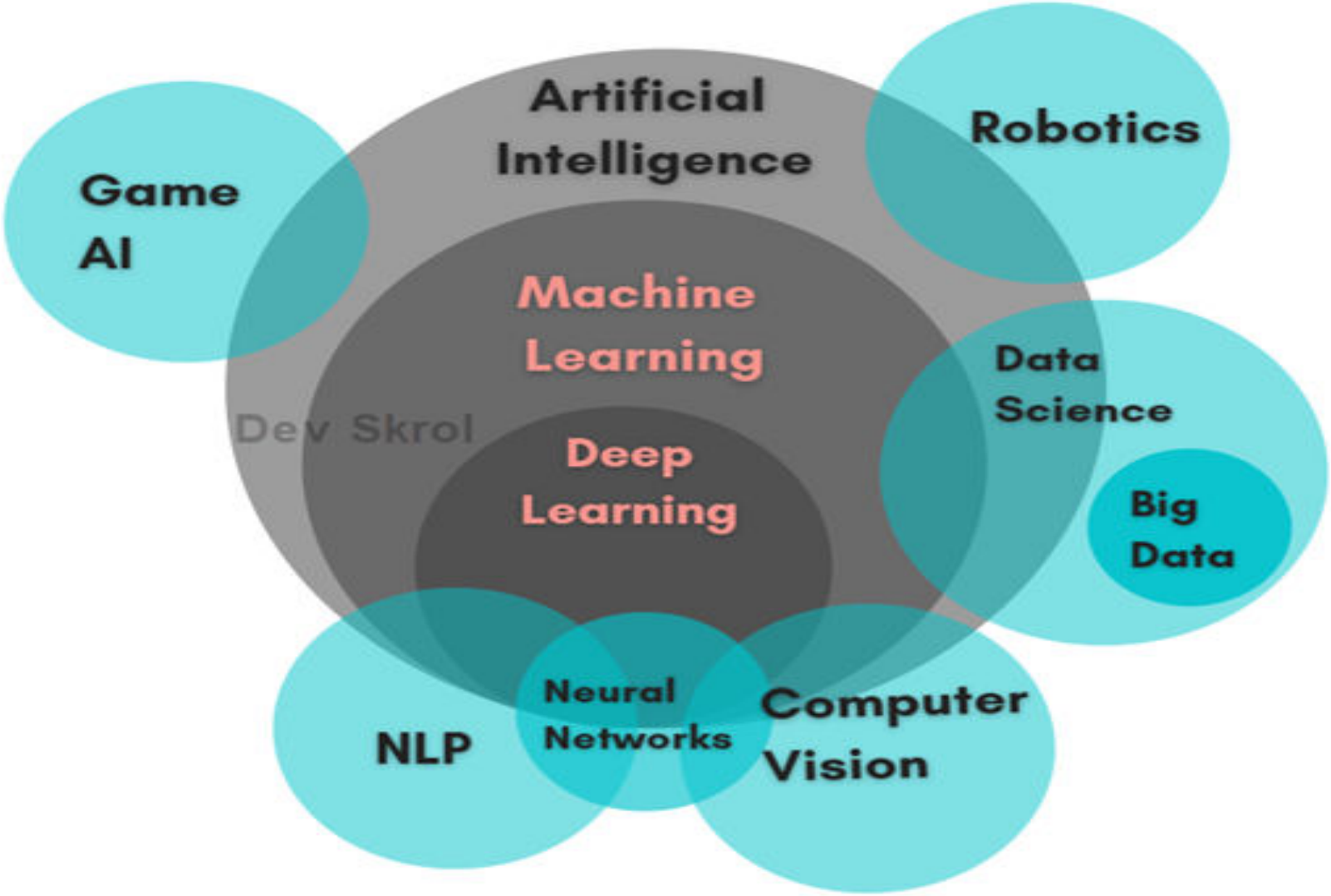
이 정 호

목차

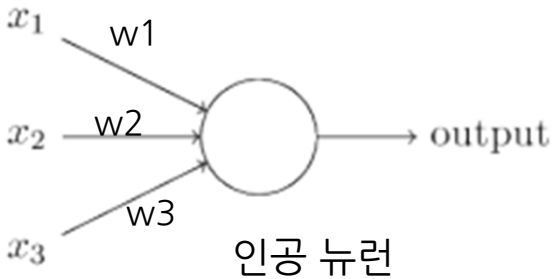
✓ 개요
✓ 심층학습 기술 효용성
✓ 심층학습 기술 문제점
✓ 심층학습 규제 고려사항
✓ 보완을 위한 기술적 방안
✓ 결론

개 요

- 인공지능 (AI) / 기계학습 / 심층학습 (DL)

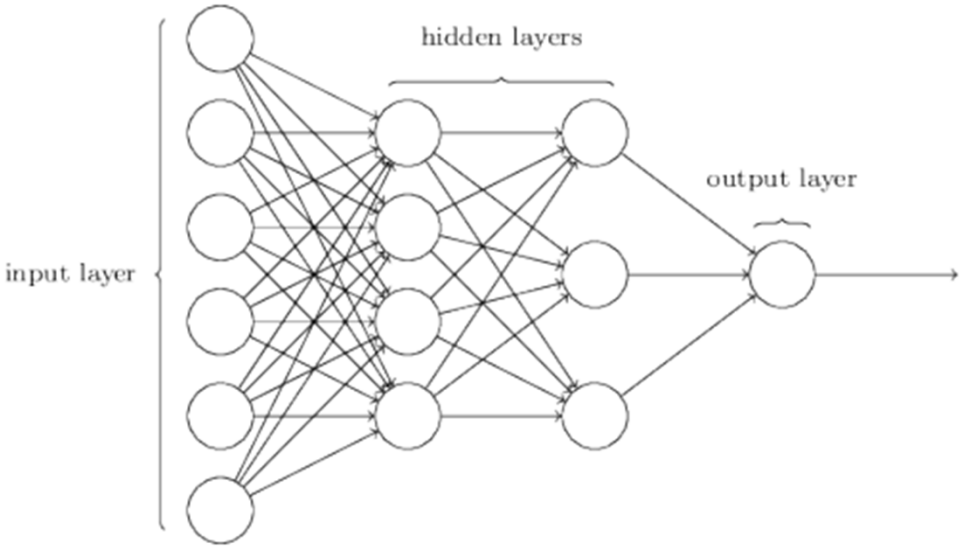


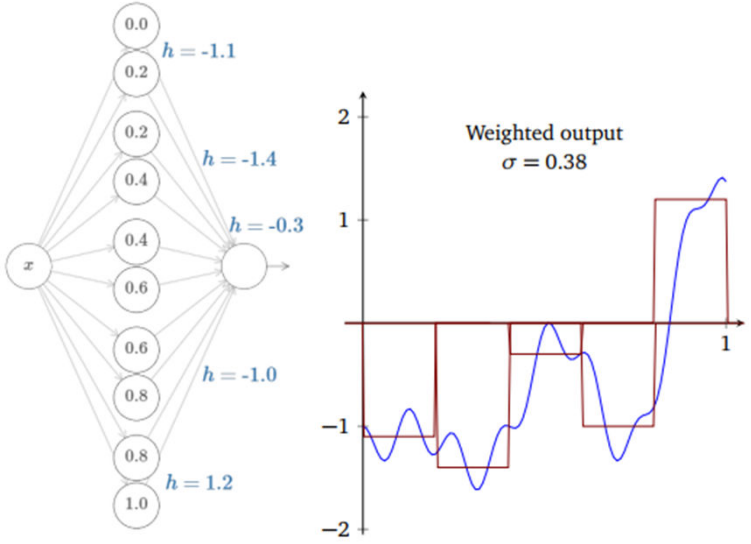
개 요



인공 뉴런

- **심층학습 이란? (실제는 아니지만 여기서는 AI = DL)**
  - 기계학습의 한 분야로서 데이터로 부터 표현을 계층적으로 학습하는 기술
  - 인공 뉴런으로 구성된 신경망을 구성
  - **심층 신경망:** 입력계층과 여러 은닉 계층, 출력계층으로 구성된 신경망
  - **학습:** 대량의 데이터셋(데이터+정답)으로 신경망의 가중치와 편향을 조절
  - **특징:** 신경망으로 **어떤 차원의 함수**라도 **모사**할 수 있음





심층학습 기술 효용성

- **원자력 산업의 인공지능 수요**
  - 기존 원전의 노후화로 인한 (안전) 수명 연장 필요
  - SMR과 같은 차세대 원전의 경제성 및 효율성 확보 필요

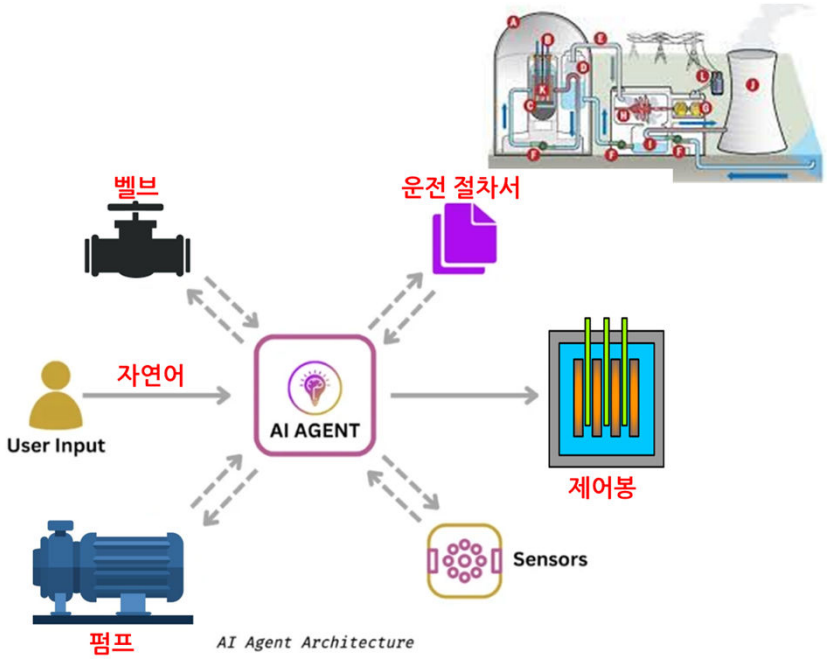
  ⇒ 심층학습 기술을 활용하여 운영 및 유지보수 비용 절감
- **심층학습 기술의 원자력분야 응용**
  - 심층학습: 인공 신경망을 기반으로 인간의 문제해결 및 의사결정 논리를 모방, 방대한 데이터를 학습하여 예측, 진단, 최적화 수행
  - 예측 정비(Predictive Maintenance), 실시간 모니터링, 핵연료 주기 최적화, 운전원 의사결정 지원 등에 활용 가능
- **고려사항**
  - 심층학습 기술의 잠재적 이점과 **내재된 위험요소(투명성, 견고성, 보안성)**는 무엇인가?
  - 안전하게 규제하고 기술적으로 극복하기 위한 전략은 무엇인가?

심층학습 기술 효용성

- **운영 효율성 제고**
  - **예측 정비(Predictive Maintenance):** 핵심 구성 요소의 고장을 사전에 예측하여 비계획적 가동 중단 방지, 설비의 수명을 연장
    - (매킨지) 유지보수 비용 최대 20% 절감, 장비 가용성 최대 50% 증가
    - 사례: 프랑스 EDF, 미국 Exelon Generation, 한국 한수원 등
  - **운전 최적화:** 노심 격자 배열, 연료 집합체 및 제어봉과 관련된 데이터를 학습하여 운전 파라미터를 최적화
    - 안전 규제 요건 내에서 연료 장전 계획을 자동 생성하고 최적화
    - 1980년대 부터 다양한 인공지능 기술이 원자로 설계에 이미 활용
    - 예: FUELCON, FUELGEN 프로젝트 등

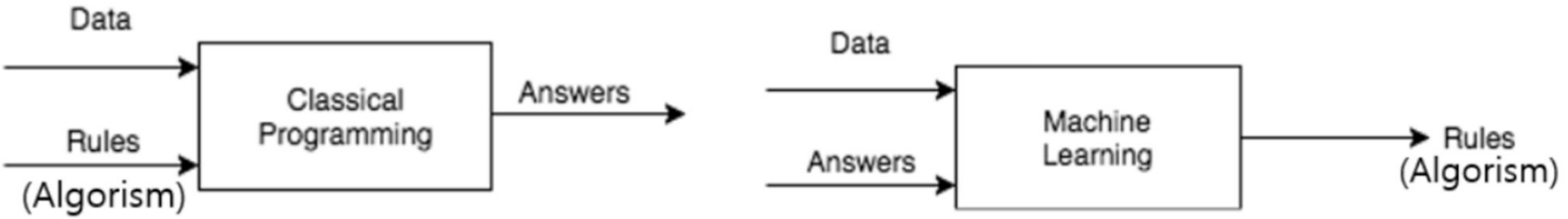
심층학습 기술 효용성

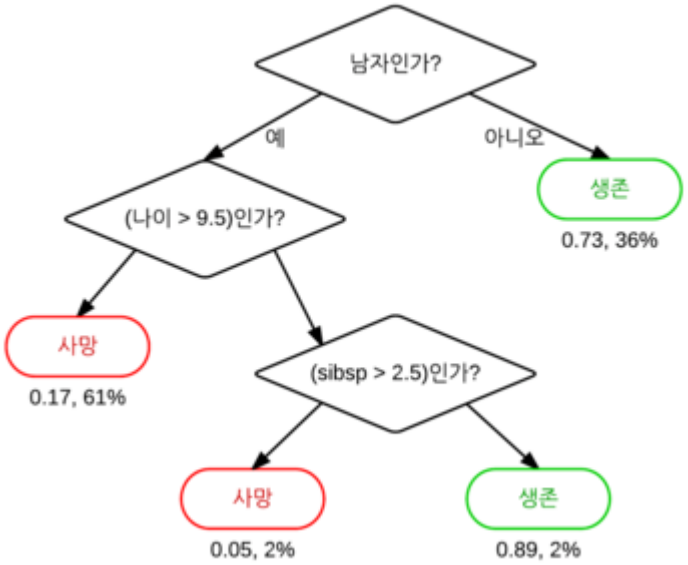
- **안전성 강화 (사고 예방)**
  - **실시간 모니터링:** 원전 설비의 실시간 데이터를 학습하여 사고 징후를 사전에 감지
    - 예: 한수원 열화상 딥러닝 기술을 이용한 전력 설비의 화재 및 고장 징후 감지
  - **운전원 의사결정 지원:** 운전 절차서, 운전 파라미터를 학습하여 운전원의 인지 부하를 줄이고 신속하고 정확한 의사결정을 지원
    - 예: KAERI 완화조치 효과 예측 및 평가 기반 운전지원시스템(PRISM)
    - 원전 자율 운전?



심층학습 기술의 문제점

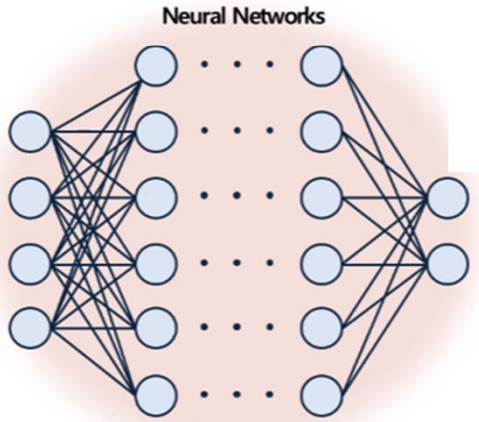
- **블랙박스 딜레마 (투명성)**
  - 심층학습: 다층 신경망+대용량 데이터, 학습결과: 신경망의 가중치와 편향
  - 한계: 추론과정이 불투명한 블랙박스로 간주





기존 (제어) SW (Decision Tree)







폐렴

심층학습 모델

심층학습 기술의 문제점

- **비결정론적 특징**
  - 심층학습: 무작위 초기화, 확률적 업데이트, 확률적 정규화
  - 두 개의 같은 신경망을 같은 데이터로 정확히 같은 절차로 훈련하더라도 결과가 완전히 일치하지 않음
- **신뢰성 문제 (Verification & Validation, Uncertainty Quantification, VVUQ)**
  - 심층학습을 위해서는 고품질의 데이터 필요(정상 운전 데이터 >> 비정상 운전데이터, 학습 데이터에 내제되어 있을지 모르는 편향이나 오류)
  - 검증(Verification): 심층신경망의 규모와 구조적 복잡성과 비선형성
  - 확인(Validation): 학습 데이터 분포를 벗어난 입력에 대한 예측 불가
  - 불확실성 정량화: 대규모 심층신경망의 가중치에 대한 계산 비용이 너무 높음

심층학습 기술의 문제점

- **모델 드리프트 (Model Drift)**
  - 운영 환경이 훈련환경과 달라지면 심층신경망 모델의 성능이 저하됨

  ⇒ 예측 정확성 저하, 잘못된 결론, 예상치 못한 행동 유발 가능 (전체 원전의 안전 여유도에 위험을 전가할 수 있음)
- **보안성**
  - (학습 / 운영) 데이터에 대한 의도적 조작, 모델 가중치 및 편향의 미세한 조작 등
  - (공급망) 심층학습 기술을 구성하는 방대한 SW 라이브러리를 활용한 공격 등

심층학습 규제 고려사항

- **안전 보증 프레임워크 구축 필요**
  - 인공지능 기술의 원자력 분야 적용을 위해서는 기술적 완성도 뿐만 아니라 조직 및 규제적 준비도 요구됨
  - IAEA: AI 기술의 안전성 및 보안성 지침 개발을 위하여 회원국의 경험을 수집, AI시스템의 사이버보안 CRP 추진
  - OECD NEA: 규제기관이 기술개발 과정 조기에 참여하는 프로세스를 개발(RegLab)
- **규제를 위한 V&V 방법론 개발 필요**
  - IEEE 권고: 딥러닝 평가 프레임워크(IEEE 2841-2022)
  - Explainable AI (XAI) 아키텍처 가이드(IEEE 2894-2024)
  - 인공지능 견고성 시험 및 평가(IEEE 3129-2023)
  - ISO/IEC 표준: AI 위험 관리 지침(ISO 23894:2023), AI 관리시스템 표준(ISO 42001:2023)

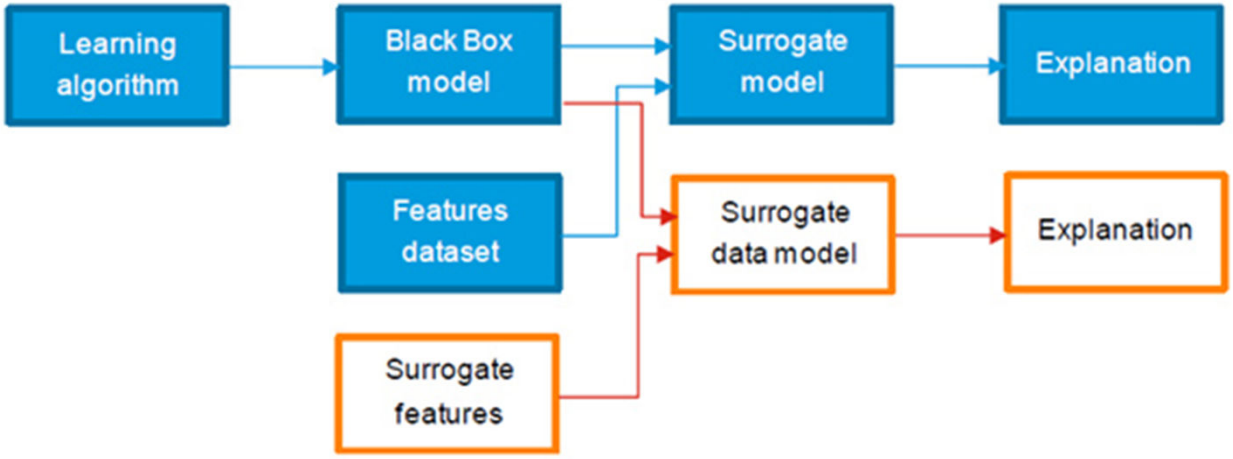
심층학습 규제 고려사항

- **신뢰모델의 패러다임 전환**
  - 신뢰의 범위 확대: 인공지능 알고리즘의 내적 무결성, 데이터 거버넌스, AI 추론 및 해석과정
  - 인공지능의 위험이 완전히 파악되기 전까지 인간 감독 유지

규제 영역	전통적 시스템	심층학습 시스템	관련 국제표준
V&V 및 신뢰성	결정론적 로직 기반 IEEE 1012/RG 1.168	VVUQ, 데이터 드리프트 및 성능 저하 감시	IEEE 2841-2022 (심층학습 평가)
안전 보증	설계기준 및 절차 준수	설명 가능성을 통한 의사 결정 과정 투명화	IEEE 2894 (XAI 아키텍처)
사이버보안	네트워크 경계 통제, 형상 관리, 접근제어	모델 견고성 확보, 데이터 조작 방지, AI 기반 위협 탐지	ISO 23894 (위험 관리)

심층학습 보완을 위한 기술적 방안

- **설명(해석) 가능한 인공지능 (Explainable AI, XAI)**
  - 심층학습 모델의 동작 방식을 해석하여 특정 데이터나 특징이 결과에 기여한 방법에 대한 통찰 제공
  - **설명용 모델 (Interpretable Models)**
    - 학습단계에서부터 AI 모델과 유사한 (단순화된) 선형 회귀, 로지스틱 회귀, 결정 트리 등 대리모델(Surrogate Model)을 만듦
    - 단점: 복잡한 문제 해결에 대한 정확도가 떨어짐



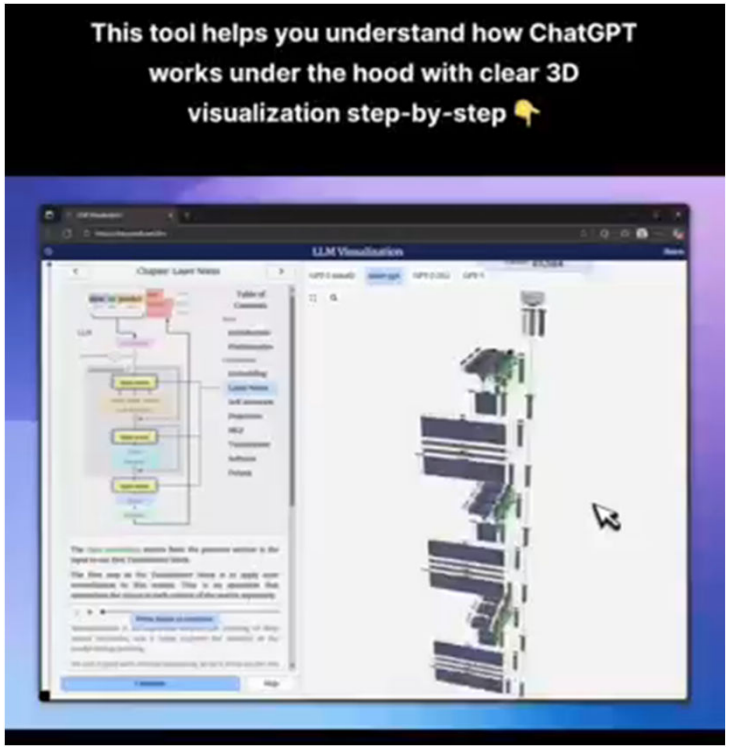
심층학습 보완을 위한 기술적 방안

- **설명(해석) 가능한 인공지능 (Explainable AI, XAI)**
  - **사후설명 기법(Post-hoc Explanation)**
    - AI 모델의 결과를 가지고 왜 그런 결과가 나왔는지를 해석
    - **LIME (Local Interpretable Model-agnostic Explanation)**
      - AI 모델 학습에 사용한 데이터를 근사하게 변형하여 어떤 현상이 발생하는지를 조사
    - **SHAP (Shapley Additive exPlanations)**
      - 각 특징이 결과에 얼마나 기여했는지를 나타내는 Shapley Value를 산출하여 예측결과를 설명
      - 장점: AI 모델의 특징을 일관성 있게 설명 가능
      - 단점: 학습하는 특징의 지수만큼 계산량 증가

심층학습 보완을 위한 기술적 방안

- **설명(해석) 가능한 인공지능 (Explainable AI, XAI)**
  - **시각화 (Visualization)**
    - 영향력 시각화 (Feature Importance Visualization)
    - 내부 작동 시각화 (Internal Mechanism Visualization)
    - 데이터 및 편향 시각화 (Data and Bias Visualization)





심층학습 보완을 위한 기술적 방안

- **Physical-informed Neural Networks (PINNs) 개발**
  - 신경망 모델에 알려진 물리 법칙을 통합하여 과학 및 공학 문제를 해결
  - 물리법칙을 신경망의 손실함수에 반영하여 유체역학, 열전도, 재료과학, 생물학 등에 응용
  - 예: 미국 아르곤 국립연구소 AI Assisted Safety Modeling and Analysis of Advanced Reactors
- **모델 견고성 및 사이버 공격 방어 기술**
  - **적대적 훈련(Adversarial Training):** 일반 훈련 데이터 외에 공격자가 조작한 적대적 예제를 포함하여 모델을 훈련
  - **공격 정규화(Regularization):** 훈련 과정에서 교란에 패널티를 부과하도록 학습 알고리즘 수정
  - **데이터 전처리:** 데이터의 미묘한 교란의 영향을 줄이도록 입력 데이터를 전처리

심층학습 보완을 위한 기술적 방안

- **심층학습 모델의 생명주기 관리**
  - 입력 데이터에 대한 범위와 조건 감지 기능 도입
  - 모델의 성능 저하 모니터링을 위한 모니터링 제어 장치 구축
  - 데이터 거버넌스를 통한 데이터의 신뢰성과 가용성 확보
  - 사이버보안을 위한 네트워크 환경 통제, 안전한 데이터 전송, HW 수준의 보안 강화 등

결 론

- **심층학습 기반 시스템 신뢰성 확보**
  - XAI, PINNs 등 심층학습 기반시스템의 견고성 확보기술 개발
- **데이터 거버넌스 및 벤치마킹 환경 구축**
  - 원전 운영 데이터 (시계열 데이터 등)을 수집/공유할 수 있도록 제도 정비
  - 시스템의 성능을 객관적으로 평가하고 검증하기 위한 방안 마련
- **인적 역량 강화**
  - AI 전문가와 원자력 엔지니어간 격차 해소, 소통 강화
  - AI 시스템에 대한 운전원 신뢰를 유지하고 인지 부하를 최소화하기 위한 방안
- **규제기관의 선제적 참여**
  - OECD NEA의 RegLab 등을 참조하여 규제 격차 해소

참고자료

[1] Accelerant. (n.d.). AI in Nuclear Plants. Retrieved from https://discoveraccelerant.com/ai-in-nuclear-plants/

[2] International Atomic Energy Agency (IAEA). (n.d.). Challenges in the Deployment of AI/ML Technologies in Nuclear Power.. Retrieved from https://www-pub.iaea.org/MTCD/publications/PDF/p15866-PUB2119_web.pdf

[3] 첨단정보기술연구소. (n.d.). AI 기반 다목적 최적화 및 보안 체계 강화.. Retrieved from https://www.cbist.or.kr/home/board/download.do?mncd=1284&fno=7866&bid=00000322&did=00005626

[4] Kubelwa, N. N. (n.d.). Enhancing Nuclear Power Production with Artificial Intelligence. IAEA Bulletin. [Article]. Retrieved from https://www.iaea.org/bulletin/enhancing-nuclear-power-production-with-artificial-intelligence

[5] techUK. (n.d.). AI-powered predictive maintenance for nuclear.. Retrieved from https://www.techuk.org/resource/ai-powered-predictive-maintenance-for-nuclear.html

[6] 한국수력원자력. (n.d.). 자동 예측진단 시스템 운영 및 기대 효과.. Retrieved from https://www.youtube.com/watch?v=Vov3sulfi_g

[7] Roumiantsev, A., & Ivanova, L. (2021). Artificial Intelligence Techniques in Nuclear Reactor Core Design and Fuel Management. Molecules, 26(3), 738. [Article]. Retrieved from https://www.mdpi.com/2411-9660/3/3/37

[8] 박지훈. (2024). 인공지능 기반 운전지원시스템(PRISM) 개발 현황. [원전계측제어심포지엄 (NuPIC) 발표 요약]. Retrieved from https://www.inpnews.kr/news/articleView.html?idxno=6103

[9] [한국원자력연구원]. (2019). AI 기반 자율 운전 알고리즘 개발 및 플랫폼 구축. [보고서 요약서]. Retrieved from https://scienceon.kisti.re.kr/srch/selectPORSrchReport.do?cn=TRKO201900027191

[10] Alabdulkarim, M., & Assi, A. (2025). AI Application in Nuclear Emergency Preparedness: Challenges and Future Outlook. Frontiers in Mechanical Engineering, 11. [Article]. Retrieved from https://www.frontiersin.org/journals/mechanical-engineering/articles/10.3389/fmech.2025.1594397/full

참고자료

[11] OECD Nuclear Energy Agency (NEA). (2025). Nuclear Innovation and Newly Emerging Disruptive Technologies.. Retrieved from https://www.oecd-nea.org/upload/docs/application/pdf/2025-11/nuclear_innovation_and_newly_emerging_disruptive_technologies__recent_initiatives_by_industry_research_laboratories_and_gove.pdf

[12] IAEA. (n.d.). How Artificial Intelligence Will Change Information and Computer Security in the Nuclear World. IAEA Bulletin. [Article]. Retrieved from https://www.iaea.org/bulletin/how-artificial-intelligence-will-change-information-and-computer-security-in-the-nuclear-world

[13] IAEA. (2022). The Role of the IAEA in the Deployment of AI for Nuclear Power.. Retrieved from https://www-pub.iaea.org/MTCD/publications/PDF/p15866-PUB2119_web.pdf

[14] OECD NEA. (n.d.). Task Force on Artificial Intelligence and Machine Learning for Scientific Computing in Nuclear Engineering.. Retrieved from https://www.oecd-nea.org/jcms/pl_77779/task-force-on-artificial-intelligence-and-machine-learning-for-scientific-computing-in-nuclear-engineering

[15] US NRC. (2024). Artificial Intelligence/Machine Learning Risk Transfer Considerations.. Retrieved from https://www.nrc.gov/docs/ML2424/ML24241A252.pdf

[16] IAEA. (2024). New Research Project on Computer Security for Nuclear AI.. Retrieved from https://www.iaea.org/newscenter/news/new-research-project-on-computer-security-for-nuclear-ai

[17] Aoun, A. F. (2024). Explainable AI Framework for False Data Injection in Nuclear Reactor Signals. [arXiv preprint]. Retrieved from https://arxiv.org/html/2403.13502v2

[18] Barbosa, M. G., & Vidal, M. C. R. (2018). Human-centered design for nuclear power plant control room. [Conference Paper]. Retrieved from https://www.semanticscholar.org/paper/Human-centered-design-for-nuclear-power-plant-room-Universit%C3%a1ria/a48280c4905132c97399e6bf73a966d9acb8503d

[19] Handa, K. (2007). Human-machine interaction in nuclear power plants. Journal of Power and Energy Engineering, 1(2). [Article]. Retrieved from https://www.researchgate.net/publication/228628218_Human-machine_interaction_in_nuclear_power_plants

[20] IAEA. (n.d.). Technical Meeting on the Safety Implications of the Use of AI in Nuclear Power Plant. [Event Page]. Retrieved from https://aiforgood.itu.int/about-us/un-ai-actions/iaea/

참고자료

[21] US NRC. (2013). Secure Development and Operational Environment (SDOE) Guidance.. Retrieved from https://www.nrc.gov/docs/ML1307/ML13073A210.pdf

[22] US NRC. (2024). AI Regulatory Gap Analysis: Review of Standards.. Retrieved from https://www.nrc.gov/docs/ML2429/ML24290A059.pdf

[23] OECD NEA. (n.d.). International RegLab Project.. Retrieved from https://www.oecd-nea.org/jcms/pl_104274/international-reglab-project

[24] OECD NEA. (2025). AI in Nuclear Operations: RegLab Workshop Convenes Regulators, Operators and Vendors.. Retrieved from https://www.oecd-nea.org/jcms/pl_111815/ai-in-nuclear-operations-reglab-workshop-convenes-regulators-operators-and-vendors

[25] OECD NEA. (n.d.). Harnessing AI: First Workshop of the New NEA Joint Project AIxpertise.. Retrieved from https://www.oecd-nea.org/jcms/pl_111870/harnessing-ai-first-workshop-of-the-new-nea-joint-project-aixpertise

[26] M., D., & J., A. (2024). Review of Deep Learning Applications in Nuclear Power Plant Control Systems. arXiv preprint. Retrieved from https://arxiv.org/html/2403.13502v2

[27] Abur, M., & Yilmaz, Y. (2024). Explainable Deep Learning for Real-time Anomaly Detection in Nuclear Power Plants. Electronics, 13(22), 4428. [Article]. Retrieved from https://www.mdpi.com/2079-9292/13/22/4428

[28] M., D., & J., A. (2024). Defense Methods Against Adversarial Attacks in Deep Learning for Nuclear Applications. [Article]. Retrieved from https://arxiv.org/html/2403.13502v2

[29] M., D., & J., A. (2024). Defense Methods Against Adversarial Attacks in Deep Learning for Nuclear Applications. Electronics, 13(22), 4428. [Article]. Retrieved from https://www.mdpi.com/2504-2289/8/1/8

[30] Argonne National Laboratory (ANL). (n.d.). AI-Assisted Safety Modeling and Analysis of Advanced Reactors. [Project Page]. Retrieved from https://www.anl.gov/ntns/aiassisted-safety-modeling-and-analysis-of-advanced-reactors

[31] Aoun, A. F. (2024). Physics-Informed Neural Networks for Reactor Transient Prediction.. Retrieved from https://inldigitallibrary.inl.gov/sites/sti/sti/Sort_151656.pdf

감사합니다.