AI기반 필수 디지털 자산 유형 분류 기법

AI-based CDA Type Identification Methodology

2026 . 5

이 원 영

INDEX

AI-based CDA Type Identification Methodology

1 연구 배경 및 필요성

AI기반 필수 디지털 자산 유형 분류 기법

1. 연구 배경 및 필요성

연구 배경

- 원자력시설의 사이버보안은, Safety and safety-related, Security, Emergency Preparedness (SSEP) 기능을 수행하는 Critical Digital Asset (CDA)를 식별하고 이를 사이버공격으로부터 보호하는 것을 목적으로 한다.
CDA는 USNRC RG 5.71 [1] 및 NEI 08-09[2]에 따라 적절히 보호되어야 하나, [1]과 [2]의 security control은 기기의 특성 및 그에 따른 Attack vector를 고려하지 않은 General requirement이다.

- 이에 따라 NEI 13-10 [3]은 NEI 10-04[4]에 따라 식별된 CDA를 A1~B3 유형으로 분류하고 이에 따라 Graded Approach를 위한 Framework를 제공한다.
CDA의 유형 분류는 기기의 특성(communication protocol, HMI characteristic, operating system, et. Cetera.)을 기준으로 수행되며 이는 사이버보안 담당자의 판단에 따라 이루어지고 있다.

1. 연구 배경 및 필요성

연구 필요성

•사이버보안 설계 담당자는 공급자 기술 문서를 통해 CDA의 특성, 기능을 파악하고 이를 기반으로 CDA 분류를 수행하여야 한다.

• 발전소 내에 존재하는 수천 개의 CDA에 대한 비정형의 방대한 공급자 도서와 기술 문서를 종류별로 모두 검토하는 데에는 많은 시간이 소요된다

• 담당자의 역량과 개개인의 판단 차이에 따라 식별 결과가 일정하지 않을 수 있으며, 인적 요소로 인한 분류 오류도 존재한다.

• CDA 유형 분류 과정에서 발생하는 검토 및 분석 역무로 인한 부담과 담당자 역량 및 인적 오류로 인해 발생하는 부담을 경감시키는 것을 목표로 본 연구는 NEI 13-10에서 제시하는 CDA 유형 분류 기준과 디지털 자산의 기술 문서, Catalog, Data sheet 등을 LLM이 분석하여 Direct CDA 유형분류를 자동화하는 모델을 제시한다.

2 연구 방법

AI기반 필수 디지털 자산 유형 분류 기법

2. 연구 방법

데이터 수집

- NEI 13-10, Appendix D에 예시로 공개된 기기의 Catalog, Technical Manual, Specification 등 기술 문서 수집
- NEI 13-10의 식별 예시를 Ground truth로 설정
- 일반화 성능 평가를 위해 원전에서 사용되는 기기와 유사한 기기의 기술 문서 수집 및 레이블링

필요 정보 정의

- NEI 13-10, NEI 10-04등 Cybersecurity Code&Standards를 분석하여 유형분류 요건을 명세
- 유형분류에 필요한 기기의 특징 및 정보를 식별

정보 추출 모델 개발

- Expert persona 부여를 통한 원자력 사이버보안 Context로 분야 한정
- LLM을 활용해 제품명, 통신 프로토콜, HMI의 존재 여부 및 설정 가능 기능 등 필요한 정보를 자동으로 추출
- PDF Plumber library를 사용

3. 연구 방법

CDA 유형 분류 모델 개발

- NEI 13-10의 유형 분류 기준을 CoT를 활용한 Decision tree로 만들어 각 Step을 LLM에 제공
- LLM은 각 Step별 판단 근거를 출력으로 제공
- 판단 과정을 돕기 위한 Few-shot 제공
- Expertpersona + Few-shot + Chain-of-Thought를 활용하여 생성한 Prompt와 기술 문서에서 추출한 정보를 토대로 CDA 유형을 분류

성능평가

- Ground truth와 비교하여 분류 정확도를 분석
- 오류 발생의 경우를 분석하여 실패 요인을 식별
- 실제 발전소의 CDA와 유사한 기기의 기술문서를 수집하여 완성된 모델의 일반화 성능 검증

3 NEI 13-10 분석

AI기반 필수 디지털 자산 유형 분류 기법

3. NEI 13-10 분석

NEI 13-10 유형분류 기준 분석

		A1	A2	A3	B1	B2	B3
software attributes		Program code (e.g. instruction-level code) cannot be altered and does not utilize or support operating system or application software	Program code (instruction-level code) is factory installed by manufacturer and cannot be altered nor can any code be injected (e.g. no buffer or heap overflow)				Program code (instruction-level code) is factory installed by manufacturer, but it may be possible to replace this program code by doing a firmware update in the field.
							If the CDA supports a USB port ('master' or 'slave') it is factory- programmed to support a specific subset of bulk-data/file- exchange methods (e.g., save or reload CDA configuration settings or load new firmware). The CDA does not support interoperation with any other form of USB object classes or devices or allow automatic/manual installation of third-party drivers for such object classes or devices.
	HMI	Changes to operational parameters or operational settings can only be implemented using maintenance and test equipment	Only operational parameters (no configuration settings) can be changed using the local, integral HMI	Operational parameters can be changed using the local, integral HMI			
			No configuration changes can be made via the integral HMI	Configuration settings can be changed using the local, integral HMI			
			The HMI has no access enforcement mechanisms	The HMI has at least one form of software access enforcement mechanism			
				Does not support multi-users and individual authentication for those users			
		Configuration changes can only be implemented by taking the device out of service	Configuration setting changes can only be made using a maintenance tool and only by taking the CDA out of service	Configuration setting changes can also be made using a maintenance tool and only by taking the CDA out of service		Configuration changes can also be made using a maintenance tool and only by taking the CDA out of service.	
							Configuration changes may also be made locally via a console port and/or USB thumb drive/memory card as well as remotely via the asynchronous serial communication
			Does not contain an externally accessible file system				Does not contain an externally accessible file system but may support bulk data extraction and configuration loading/saving via the USB/memory card
			Firmware updates not supported/not possible by the hardware design				CDA supports firmware update/replacement with removal of the CDA from the service and use of special tools and software
			Contain vendor software that performs/supports a pre-defined set of features and functions and supports no ability to add or remove software	Only contains vendor's software that performs/supports a pre- defined set of features and functions and supports no ability to add or remove software			
	Communication		Contains no communication software functionality		The CDA uses an industrial protocol using poll-response based message exchanges over an asynchronous serial communications channel. Communication functionality of the CDA are limited to information or data extraction and do not support the capability for control execution, manipulation of CDA I/O or sending parameters or data to the CDA.	The CDA uses an industrial protocol using poll-response based message exchanges over an asynchronous serial communications channel. Communication functionality of the CDA can be adjusted and altered by the user and may include reading and writing data from and to the CDA to fetch values, change/set parameters, execution of pre-configured control functions and manipulation of CDA process control outputs.	
					Communication functions do not allow for modification of the configuration of the CDA or for making program changes to the CDA.		The functionality and configuration of the CDA can also be altered via these communication links using software tools (possibly vendor- proprietary) specifically designed for that purpose.
							The asynchronous communications capability does not support modification of code, instructions, or code injection to the CDA.

3. NEI 13-10 분석

NEI 13-10 유형분류 기준 분석

		A1	A2	A3	B1	B2	B3
hardware attributes		Device does not support any sort of event logging	CDA does not perform audit/event logging of user activities or communication activities or local runtime events.				
		Device does not support application or 3rd party software					
			The CDA does not suppler a local console port or command line interpreter functionality				The CDA has a local, special- purpose communications interface (a.k.a. a console port), typically a low-speed, asynchronous, EIA-232 compatible, that is used to enable user interaction with a device's integral command-line interpreter (e.g., a "shell" or "command prompt") via an ASCII 'dumb terminal' or a computer/program emulating a dumb terminal
		Device includes PROM, RAM, EEPROM and possibly integrated components (e.g. FPGA) with factory-configurable firmware and functionality	Contain PROM, RAM, EEPROM, and possibly integrated components (e.g., FPGA) that include factory- configurable functionality and factory-configurable firmware		PROM, RAM, EEPROM and possibly integrated components (e.g., FPGA) that include factory-configurable functionality and factory- configurable firmware.		
			May contain bulk storage for data accumulation purposes but provides no external access to that bulk storage				May contain bulk storage for data accumulation purposes and for configuration setting storage and May support external access to that bulk storage.
	HMI	Device has no remote or local, integral HMI (but may have local display-only indicators)	Has a minimal-functionality				
			Local access only				
				May employ a physical access protection mechanism such as a key or fob			
			Contains a maintenance and configuration port but no other peripherals, interfaces, or ports		Contains a maintenance and configuration port as well as one or more asynchronous communication ports but no other peripherals, interfaces or ports		Contains a console port and one or more non-Ethernet serial communication ports (synchronous or asynchronous)
							May support a restricted functionality USB port and/or memory card slot for bulk data retrieval and configuration exporting and restoration but no other peripherals, interfaces or ports
		Device has no communications hardware/software but may have interfaces to external devices/systems using analog/contact/pulse I/O signals	May support an interface to external devices/systems implemented using analog, contact, pulse process control I/O signals				
		Device has no peripherals, interfaces or ports (e.g. media access, serial, etc.)					
Location		Protected Area (PA) or Vital Area (VA)					
Information Calssification		CDA contains plant process data not classified as security-related or Safeguards Information (SGI)					
Plant Design / Maintenance		Removal from service can only be done locally at the CDA					
If the CDA contains peripherals, interfaces, or ports beyond those allowed by the class criteria, the CDAs can meet this criteria by physically disabling the peripheral, interfaces or ports in a manner that prevents restoration, reactivation or bypass.							

- 의미없이 비어 있는 셀, 동일 기준의 반복 작성, 모든 등급에 기준이 동일한 경우 등 다수 불필요 정보 존재
- Location, Information classification, Plant design, maintenance는 기기의 technical document로 판단이 불가능한 사항임
- 상기 사항을 포함하여 대분류 등 LLM 입장에서 Noise일 수 있는 내용을 삭제

3. NEI 13-10 분석

NEI 13-10 유형분류 기준 분석

		A1	A2	A3	B1	B2	B3
software attributes		Program code (e.g. instruction-level code) cannot be altered and does not utilize or support operating system or application software	Program code (instruction-level code) is factory installed by manufacturer and cannot be altered nor can any code be injected (e.g. no buffer or heap overflow)				Program code (instruction-level code) is factory installed by manufacturer, but it may be possible to replace this program code by doing a firmware update in the field.
							If the CDA supports a USB port ('master' or 'slave') it is factory- programmed to support a specific subset of bulk-data/file- exchange methods (e.g., save or reload CDA configuration settings or load new firmware). The CDA does not support interoperation with any other form of USB object classes or devices or allow automatic/manual installation of third-party drivers for such object classes or devices.
	HMI	Changes to operational parameters or operational settings can only be implemented using maintenance and test equipment	Only operational parameters (no configuration settings) can be changed using the local, integral HMI	Operational parameters can be changed using the local, integral HMI			
			No configuration changes can be made via the integral HMI	Configuration settings can be changed using the local, integral HMI			
			The HMI has no access enforcement mechanisms	The HMI has at least one form of software access enforcement mechanism			
				Does not support multi-users and individual authentication for those users			
		Configuration changes can only be implemented by taking the device out of service	Configuration setting changes can only be made using a maintenance tool and only by taking the CDA out of service	Configuration setting changes can also be made using a maintenance tool and only by taking the CDA out of service		Configuration changes can also be made using a maintenance tool and only by taking the CDA out of service.	
							Configuration changes may also be made locally via a console port and/or USB thumb drive/memory card as well as remotely via the asynchronous serial communication
			Does not contain an externally accessible file system				Does not contain an externally accessible file system but may support bulk data extraction and configuration loading/saving via the USB/memory card
			Firmware updates not supported/not possible by the hardware design				CDA supports firmware update/replacement with removal of the CDA from the service and use of special tools and software
			Contain vendor software that performs/supports a pre-defined set of features and functions and supports no ability to add or remove software	Only contains vendor's software that performs/supports a pre- defined set of features and functions and supports no ability to add or remove software			
	Communication		Contains no communication software functionality		The CDA uses an industrial protocol using poll-response based message exchanges over an asynchronous serial communications channel. Communication functionality of the CDA are limited to information or data extraction and do not support the capability for control execution, manipulation of CDA I/O or sending parameters or data to the CDA.	The CDA uses an industrial protocol using poll-response based message exchanges over an asynchronous serial communications channel. Communication functionality of the CDA can be adjusted and altered by the user and may include reading and writing data from and to the CDA to fetch values, change/set parameters, execution of pre-configured control functions and manipulation of CDA process control outputs.	
					Communication functions do not allow for modification of the configuration of the CDA or for making program changes to the CDA.		The functionality and configuration of the CDA can also be altered via these communication links using software tools (possibly vendor- proprietary) specifically designed for that purpose.
							The asynchronous communications capability does not support modification of code, instructions, or code injection to the CDA.

3. NEI 13-10 분석

NEI 13-10 유형분류 기준 분석

		A1	A2	A3	B1	B2	B3
hardware attributes	HMI	Device does not support any sort of event logging	CDA does not perform audit/event logging of user activities or communication activities or local runtime events.				
		Device does not support application or 3rd party software					
			The CDA does not suppler a local console port or command line interpreter functionality				The CDA has a local, special- purpose communications interface (a.k.a. a console port), typically a low-speed, asynchronous, EIA-232 compatible, that is used to enable user interaction with a device's integral command-line interpreter (e.g., a "shell" or "command prompt") via an ASCII 'dumb terminal' or a computer/program emulating a dumb terminal
		Device includes PROM, RAM, EEPROM and possibly integrated components (e.g. FPGA) with factory-configurable firmware and functionality	Contain PROM, RAM, EEPROM, and possibly integrated components (e.g., FPGA) that include factory- configurable functionality and factory-configurable firmware		PROM, RAM, EEPROM and possibly integrated components (e.g., FPGA) that include factory-configurable functionality and factory- configurable firmware.		
			May contain bulk storage for data accumulation purposes but provides no external access to that bulk storage				May contain bulk storage for data accumulation purposes and for configuration setting storage and May support external access to that bulk storage.
		Device has no remote or local, integral HMI (but may have local display-only indicators)	Has a minimal-functionality				
			Local access only				
				May employ a physical access protection mechanism such as a key or fob			
			Contains a maintenance and configuration port but no other peripherals, interfaces, or ports		Contains a maintenance and configuration port as well as one or more asynchronous communication ports but no other peripherals, interfaces or ports		Contains a console port and one or more non-Ethernet serial communication ports (synchronous or asynchronous)
							May support a restricted functionality USB port and/or memory card slot for bulk data retrieval and configuration exporting and restoration but no other peripherals, interfaces or ports
		Device has no communications hardware/software but may have interfaces to external devices/systems using analog/contact/pulse I/O signals	May support an interface to external devices/systems implemented using analog, contact, pulse process control I/O signals				
		Device has no peripherals, interfaces or ports (e.g. media access, serial, etc.)					
Location		Protected Area (PA) or Vital Area (VA)					
Information Calssification		CDA contains plant process data not classified as security-related or Safeguards Information (SGI)					
Plant Design / Maintenance		Removal from service can only be done locally at the CDA					

- 각 등급을 특정할 수 있는 조건만 남기고 모두 삭제 (붉은 음영의 셀)
- 데이터 품질의 정량화를 위해 Prompt Signal-to-Noise Ratio (PSNR)로 지표화
  - Signal (S)은 유효한 셀의 개수, Noise (N)은 불필요한 셀의 개수
  - 불필요셀은 구조적인 요소, Out-of-context 정보, Non-discriminative 요건, 의미 중복, 목적 이탈 요건 제거
  - PSNR = S/N
- NEI 13-10의 조건을 그대로 옮긴 raw data는 0.88로 잡음의 비율이 상대적으로 높게 나타남
- 개선 후 PSNR은 1.76로 유의미한 개선을 보임

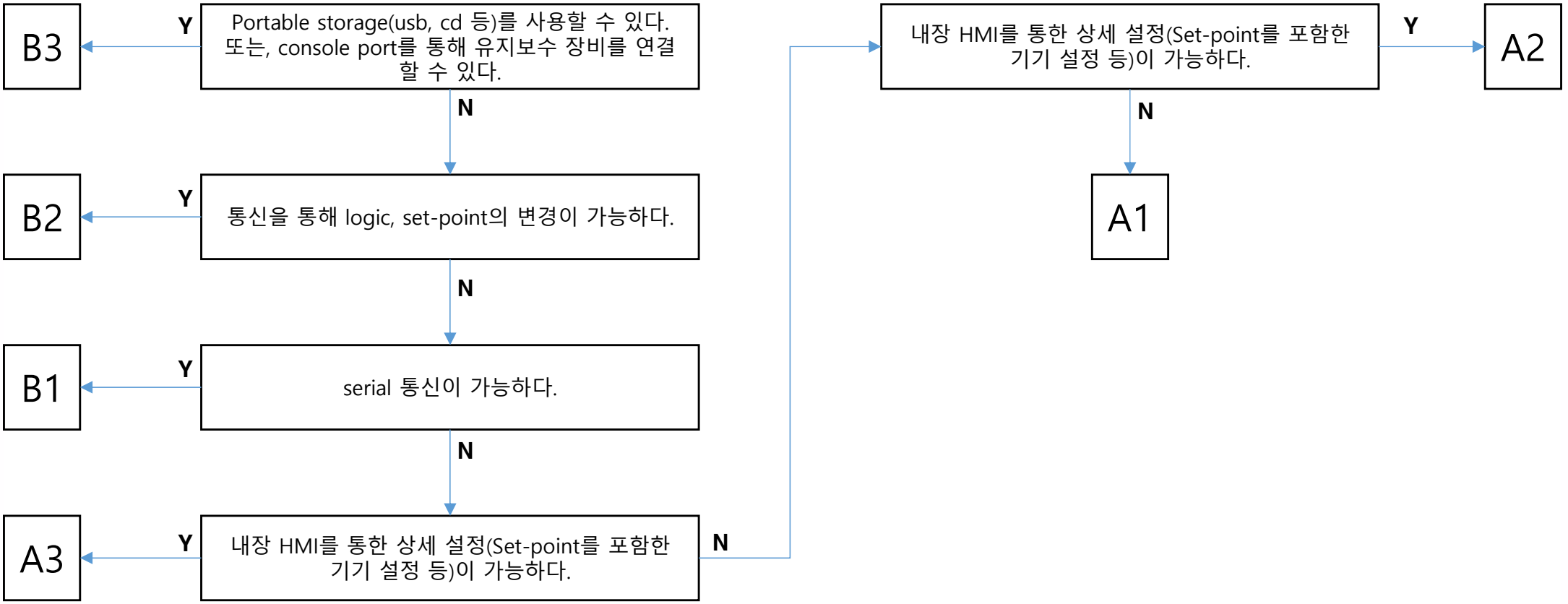
3. NEI 13-10 분석

NEI 13-10 유형분류 기준 분석

A1	A2	A3	B1	B2	B3
Changes to operational parameters or operational settings can only be implemented using maintenance and test equipment	Only operational parameters (no configuration settings) can be changed using the local, integral HMI	Operational parameters can be changed using the local, integral HMI			
	No configuration changes can be made via the integral HMI	Configuration settings can be changed using the local, integral HMI			
	The HMI has no access enforcement mechanisms	The HMI has at least one form of software access enforcement mechanism			
					Configuration changes may also be made locally via a console port and/or USB thumb drive/memory card as well as remotely via the asynchronous serial communication
	Does not contain an externally accessible file system				Does not contain an externally accessible file system but may support bulk data extraction and configuration loading/saving via the USB/memory card
	Firmware updates not supported/not possible by the hardware design				CDA supports firmware update/replacement with removal of the CDA from the service and use of special tools and software
	Contains no communication software functionality		The CDA uses an industrial protocol using poll-response based message exchanges over an asynchronous serial communications channel. Communication functionality of the CDA are limited to information or data extraction and do not support the capability for control execution, manipulation of CDA I/O or sending parameters or data to the CDA.	The CDA uses an industrial protocol using poll-response based message exchanges over an asynchronous serial communications channel. Communication functionality of the CDA can be adjusted and altered by the user and may include reading and writing data from and to the CDA to fetch values, change/set parameters, execution of pre-configured control functions and manipulation of CDA process control outputs.	
			Communication functions do not allow for modification of the configuration of the CDA or for making program changes to the CDA.		The functionality and configuration of the CDA can also be altered via these communication links using software tools (possibly vendor- proprietary) specifically designed for that purpose.
Device does not support any sort of event logging	CDA does not perform audit/event logging of user activities or communication activities or local runtime events.				
	The CDA does not suppler a local console port or command line interpreter functionality				The CDA has a local, special- purpose communications interface (a.k.a. a console port), typically a low-speed, asynchronous, EIA-232 compatible, that is used to enable user interaction with a device's integral command-line interpreter (e.g., a "shell" or "command prompt") via an ASCII 'dumb terminal' or a computer/program emulating a dumb terminal
	May contain bulk storage for data accumulation purposes but provides no external access to that bulk storage				May contain bulk storage for data accumulation purposes and for configuration setting storage and May support external access to that bulk storage.
Device has no remote or local, integral HMI (but may have local display-only indicators)	Has a minimal-functionality				
	Contains a maintenance and configuration port but no other peripherals, interfaces, or ports		Contains a maintenance and configuration port as well as one or more asynchronous communication ports but no other peripherals, interfaces or ports		Contains a console port and one or more non-Ethernet serial communication ports (synchronous or asynchronous)
Device has no communications hardware/software but may have interfaces to external devices/systems using analog/contact/pulse I/O signals	May support an interface to external devices/systems implemented using analog, contact, pulse process control I/O signals				
Device has no peripherals, interfaces or ports (e.g. media access, serial, etc.)					

3. NEI 13-10 분석

NEI 13-10 유형분류 기준 분석



4 모델 설계 및 실험

AI기반 필수 디지털 자산 유형 분류 기법

4. 모델 설계 및 실험

도출된 분류 기준을 LLM에 입력

[Level B3] (최상위 등급)
* Portable storage(usb, cd 등)를 사용할 수 있다.
* 또는, console port를 통해 유지보수 장비를 연결할 수 있다.

[Level B2]
* (B3가 아니면서) 통신을 통해 logic, set-point의 변경이 가능하다.

[Level B1]
* (B3, B2가 아니면서) serial 통신이 가능하다.

[Level A3]
* (B 등급이 아니면서) 내장 HMI를 통한 상세 설정(Set-point를 포함한 기기 설정 등)이 가능하다.

[Level A2]
* (B 등급, A3가 아니면서) 내장 HMI를 통한 set-point 설정이 가능하다.

[Level A1] (최하위 등급)
* 상기 요건(A2~B3) 중 어느 것에도 해당하지 않는 기타 디지털 기기.

4. 모델 설계 및 실험

LLM에 분류 조건 제공

```
# 등급의 상하 관계 정의
hierarchy_rule = """
[등급 결정 규칙 (High-Water Mark)]
1. [분류 기준표]의 모든 Level(B3, B2...)의 요건을 문서와 비교하여
   문서가 충족하는 모든 Level을 내부적으로 식별할 것.
2. 등급의 상하 관계(Hierarchy)는 다음과 같다:
   (낮음) A1 < A2 < A3 < B1 < B2 < B3 (높음)
3. 당신이 식별한 모든 Level 중, 이 상하 관계에 따라
   가장 높은(상위의) 등급 단 하나를 이 자산의 'final_grade'로
   선택하여야 한다.
4. 예시: 만약 문서에서 'serial 통신(B1)'과 'USB 포트 사용(B3)' 기능
   이 모두 확인된다면, B3가 B1보다 높으므로 'final_grade'는 'B3'가
   된다.
5. 만약 B3, B2, B1, A3, A2의 어떤 요건도 충족하지 않으면,
   'final_grade'는 'A1 ' 이 된다.
"""
```

```
[지시 사항]
1. 위의 [등급 결정 규칙 (High-Water Mark)]을 반드시 따라야 한
   다.
2. 결정된 'final_grade' (가장 높은 등급)를 반환하여야 한다.
3. 해당 'final_grade'를 결정하게 된 가장 결정적인 근거(rationale)
   를 [분석 대상 문서]에서 반드시 인용하여 제시하여야 한다.
   (만약 A1으로 결정되었다면, "상위 등급(B3~A2)의 요건을 찾을
   수 없음"으로 기재할 것.)
4. 해당 'final_grade' 결정에 대한 당신의 확신도
   (confidence_percent)를 0에서 100 사이의 숫자로 반환할 것.
   (문서의 근거가 명확할수록 100에 가깝습니다.)
5. 결과를 반드시 다음의 JSON 형식으로만 반환여야 한다.
   다른 설명이나 텍스트는 절대 포함하지 않는다.
```

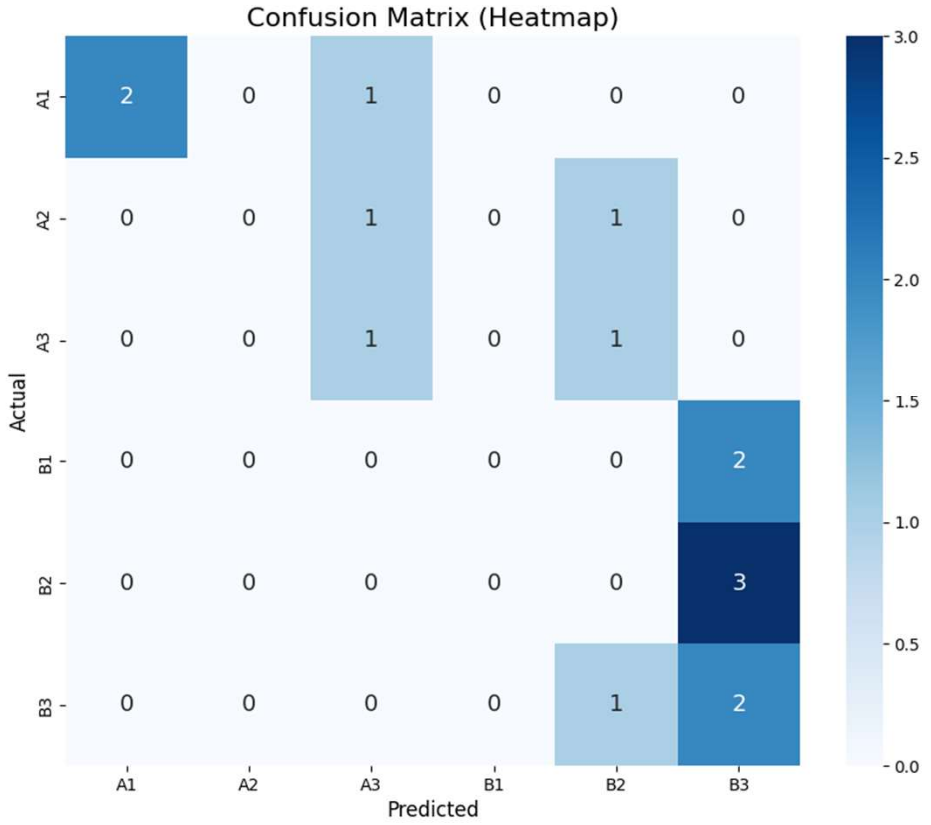
NEI 13-10의 CDA 등급 중 B3가 가장 Attack vector가 많음.

조치 사항이 가장 많은 B3의 분류가 중요하며, 식별 로직에서도 High-water Mark 규칙을 준수하도록 하였음.

4. 모델 설계 및 실험

실험 결과

	Precision	Recall	F1-score	Support
A1	1.000	0.667	0.800	3
A2	0.000	0.000	0.000	2
A3	0.333	0.500	0.400	2
B1	0.000	0.000	0.000	2
B2	0.000	0.000	0.000	3
B3	0.286	0.667	0.400	3
Accuracy	**0.333**			15
Macro avg	0.270	0.306	0.267	15
Weighted avg	0.302	0.333	0.293	15



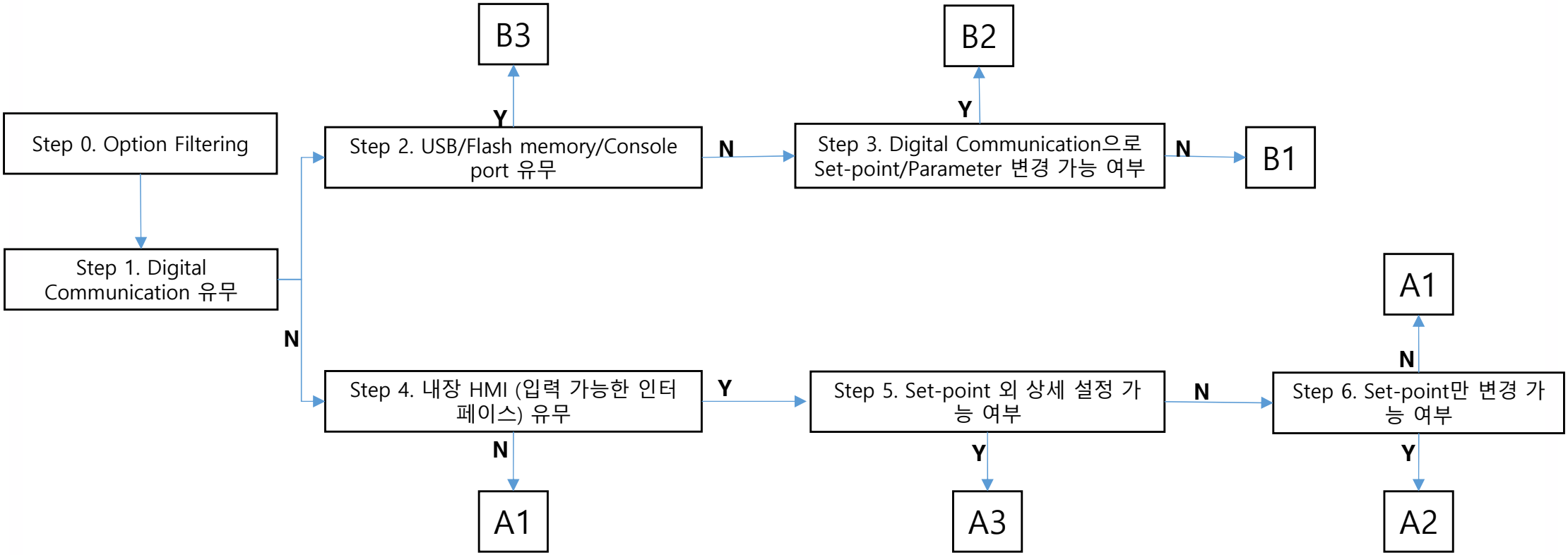
4. 모델 설계 및 실험

실험 결과 분석

- A1~3를 B2로 분류하는 경향이 보임
  - A 타입과 B 타입의 가장 큰 분기점인 Digital Communication (Interface with communication protocol)을 기술문서에서 판독하여 타입 판정에 들어가기 전에 분류하는 로직 설정
- Technical Document 내의 Optional function을 실제 기기의 특성이라고 오판하는 경우
  - File 명에 적힌 model name을 기준으로 문서를 분석
  - Option을 추가해야만 수행 가능한 기능을 식별하여 해당 기능을 판별 기준에서 제외함
- 명확한 분류 기준의 전달이 필요
  - 분류 기준이 확인됨에도 오판하는 경우에 대하여 대응이 필요
  - Few-shot (예시 6개 제공) learning을 통하여 정확도 향상 [5]
  - Few-shot과 함께 Chain-of-Thought의 제공으로 사고의 흐름을 LLM이 학습할 수 있도록 함 [6]
  - Persona prompting을 적용하여 답변의 품질, 정확성, 깊이를 향상 [7]
  - 어떤 종류의 출력을 해야 하는지 명확히 명세하여 JSON 형태로 return하게 함 [8]
- LLM이 판단에 실패한 이유를 명확히 알아야 함
  - Decision Tree를 명확히 설계하여, 각 단계에서 판단의 근거를 제공하도록 함
  - 판단에 대해 확신도를 1~100으로 표현하게 함

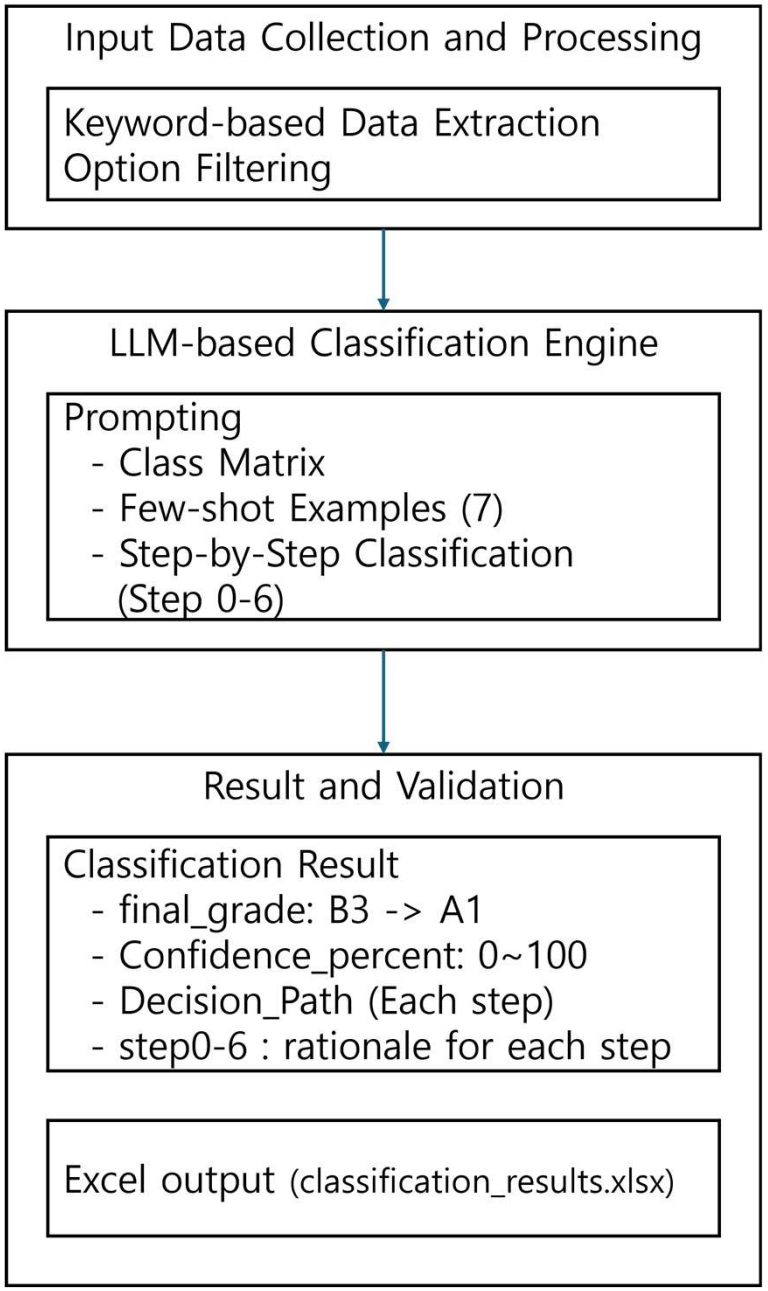
4. 모델 설계 및 실험

NEI 13-10 유형분류 기준 개선



4. 모델 설계 및 실험

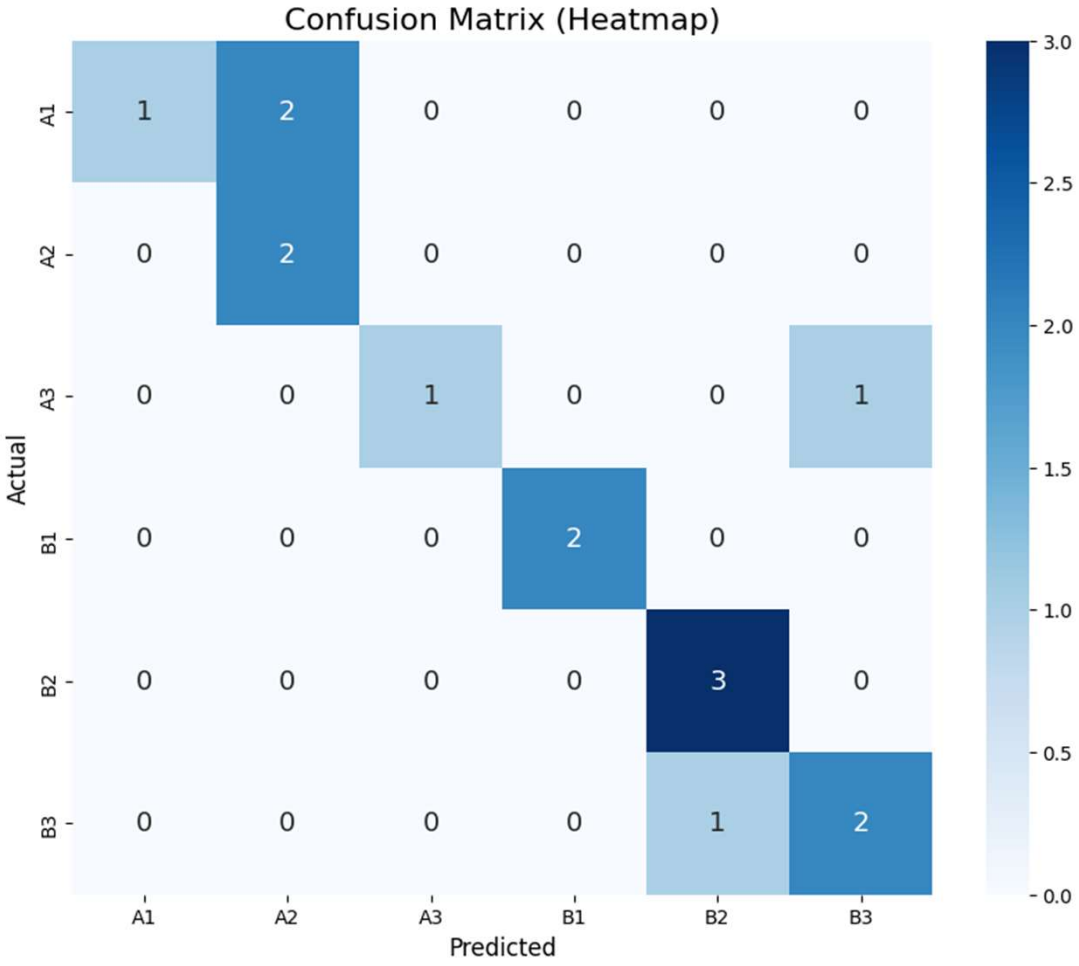
NEI 13-10 유형분류 기준 개선



4. 모델 설계 및 실험

실험 결과

	Precision	Recall	F1-score	Support
A1	1.000	1.000	1.000	3
A2	1.000	0.500	0.667	2
A3	0.667	1.000	0.800	2
B1	1.000	1.000	1.000	2
B2	0.667	0.667	0.667	3
B3	0.667	0.667	0.667	3
Accuracy		**0.800**		15
Macro avg	0.833	0.806	0.800	15
Weighted avg	0.822	0.800	0.796	15



4. 모델 설계 및 실험

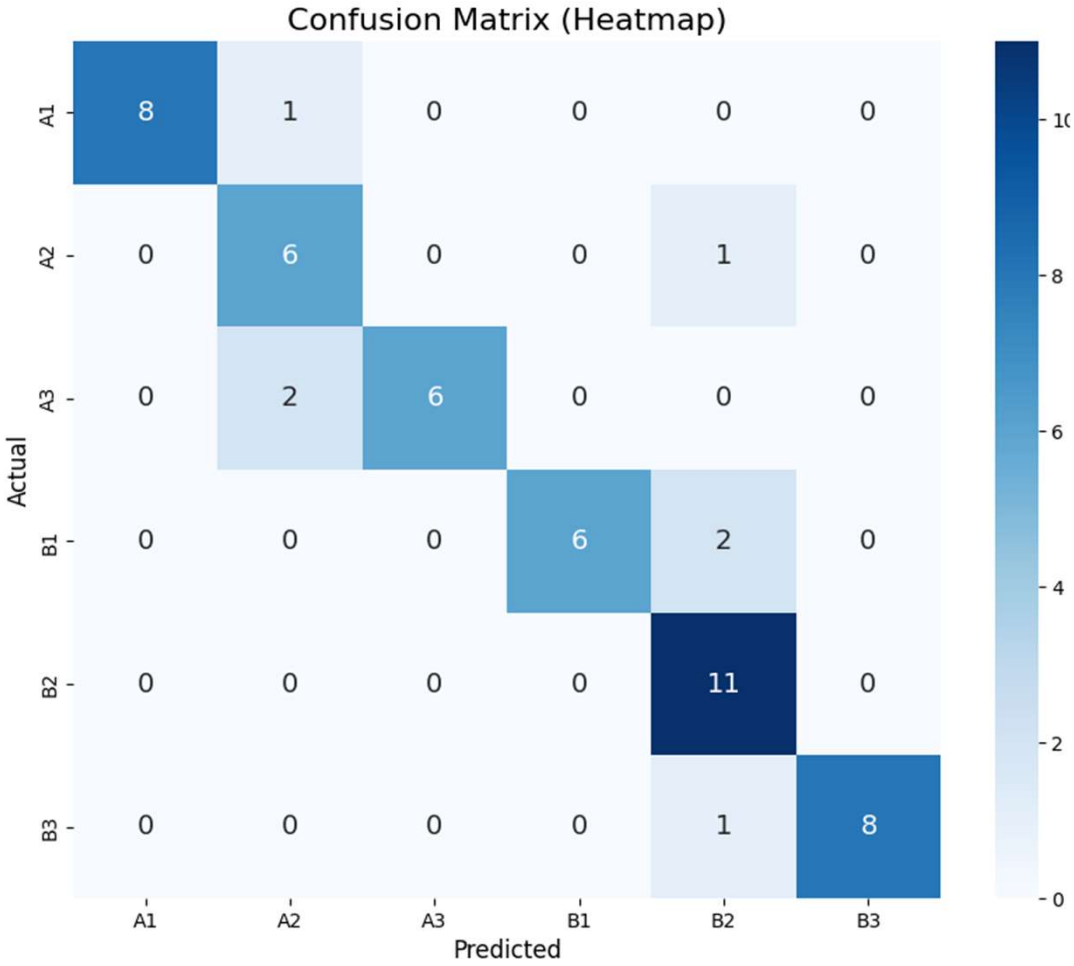
모델 오류 분석

- BK6-M (A2를 A3로 오판)
  - 기술 문서 내 '측정값 보정 설정 (bias)'이라는 텍스트를 모델이 상세 설정 (Calibration)기능으로 해석하여 A3로 오판
  - 'bias'라는 용어만으로는 Setpoint 조정(A2)인지 Calibration(A3)인지 의미적 경계가 모호하여 발생한 오류로 판단
- KH300AG (B2를 B3로 오판)
  - USB flash drive가 옵션(Option)으로 제공되나, 모델이 이를 기본 기능(Standard)으로 오인하여 B3로 과대평가
  - 카탈로그 내의 복잡한 옵션 표기 방식을 완벽히 필터링하지 못하고 기본사양으로 인식하여 오류가 발생.
- LTC0385 (B3를 B2로 오판)
  - 제조사 고유 통신 프로토콜인 Bilinx가 기본 기능임에도 불구하고, 모델이 이를 옵션 사양으로 잘못 판단하여 STEP 0에서 필터링
  - 통신을 통한 펌웨어 업데이트 기능을 인식하지 못하고 B2로 판단

4. 모델 설계 및 실험

실험 결과

	Precision	Recall	F1-score	Support
A1	1.000	0.899	0.941	9
A2	0.667	0.857	0.750	7
A3	1.000	0.750	0.857	8
B1	1.000	0.750	0.857	8
B2	0.733	1.000	0.846	11
B3	1.000	0.899	0.941	9
Accuracy		**0.865**		52
Macro avg	0.900	0.856	0.865	52
Weighted avg	0.899	0.865	0.869	52



4. 모델 설계 및 실험

Contribution

- 원자력 사이버보안이라는 특수 분야의 규제 기준인 NEI 13-10을 분석하여 LLM이 요건을 인식하기에 용이하도록 정제, 이를 활용해 LLM이 CDA 유형분류를 수행할 수 있는 토대를 마련
- Decision Tree 기반 CoT, 전문가 페르소나 등을 활용한 복합적인 프롬프트 엔지니어링 기법을 적용하여 모델의 성능 향상
- 원자력 사이버보안이라는 특수 전문 도메인의 복잡한 규제(NEI 13-10)에 따른 유형 분류를 LLM을 통해 자동화할 수 있는 실증적 가능성을 제시.

Limitation

- 단순히 분류의 기준이 되는 키워드만 인식. 문맥에 따른 추론이 불가.
- 여러 모델의 사양이나 옵션이 문서에 혼재된 경우 옵션 필터링에 실패하는 사례 존재
- Zero-tolerance가 요구되는 원자력 분야에서 86.5%의 성능은 실사용에는 부족함

5 참고 문헌

AI기반 필수 디지털 자산 유형 분류 기법

5. 참고 문헌

[1] U.S. Nuclear Regulatory Commission, *Regulatory Guide 5.71: Cyber Security Programs for Nuclear Facilities*, Rev.0, Jan. 2010.

[2] Nuclear Energy Institute, *NEI 08-09: Cyber Security Plan for Nuclear Power Reactors*, Rev.6, Apr. 2010.

[3] Nuclear Energy Institute, *NEI 13-10: Cyber Security Control Assessments*, Rev.0, 2013.

[4] Nuclear Energy Institute, *NEI 10-04: Identifying Systems and Assets Subject to the Cyber Security Rule*, Rev.1, Apr. 2010.

[5] Brown, T. B., et al. Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems (NeurIPS), 33. 2020.

[6] Wei, J., et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. Advances in Neural Information Processing Systems (NeurIPS), 35. 2022.

[7] Xu, B., Yang, A., Lin, J., Wang, Q., Zhou, C., Zhang, Y., & Mao, Z. ExpertPrompting: Instructing Large Language Models to be Distinguished Experts. arXiv:2305.14688. 2023.

[8] Min, S., et al. Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?. arXiv preprint arXiv:2202.12837. 2022.

THANK YOU

newpower, newstandard

우리의 기술이 새로운 세계기준입니다.

KEPCO E&C