ITPE 정보관리/컴퓨터시스템응용기술사 1위 교재
강의교재/서브노트 심화 겸용(강서교재)

[Domain 12]

AI/ML 이론과 전략 II (법 지침 가이드 표준)

차세대 IT 리더스 유니버시티
IT Leaders University Coursework

ICT 기술사 전문가과정
PMP 감리사 교육센터

기술사 강 정 배

목차

Chapter I. 인공지능 도메인 법 지침

회차	교시	도메인	문제
137	1	AI	6. AI 거버넌스(Artificial Intelligence Governance)를 설명하시오
137	1	AI	7. 트랜스포머(Transformer)와 MoE(Mixture of Experts)를 설명하시오.
137	1	AI	8. AI 신뢰성 검인증 제도(CAT)를 설명하시오
137	2	AI	3. MCP(Model Context Protocol)를 이용한 인공지능 서비스 구축시 보안 취약점을 설명하고 대응방안을 제시하시오
137	2	AI	4. "공공부문 초거대 AI 도입, 활용 가이드라인 2.0"에 대하여 다음을 설명하시오 가. 초거대 AI의 개념과 구성요소 나. 초거대 AI의 기술요소 다. 초거대 AI의 도입절차

회차	교시	도메인	문제
136	4	AI	1. 국내 인공지능(AI) 윤리기준과 생성형 AI에 대하여 다음을 설명하시오. 가. 3대 기본 원칙과 10대 핵심요건(과기정통부 인공지능 윤리기준) 나. 인공지능 윤리 관점에서 생성형 AI의 역기능 요소
136	4	AI	5. 대형언어모델(LIM, Large Language Model)의 활용이 급격히 증가함에 따라 그와 관련된 보안 위협이 새롭게 대두되고 있다. OWASP에서는 2025년 버전의 LLM 애플리케이션을 위한 ToP 10 보안 위협 목록(OWASP LLM, OWASP Top 10 for LLM Application 2025)을 발표하여 LLM 기반 시스템의 안전한 개발과 운영을 위한 기준을 제시하고 있다. 이와 관련하여 다음을 설명하시오. 가. OWASP LLM이 제시된 배경과 주요 특징 나. OWASP LLM에서 제시된 주요 보안 위협 다. 기업이 LLM 애플리케이션을 설계, 운영할 때 OWASP LLM을 활용한 보안 대응방안

(중요) 법 지침 가이드처럼, 정답이 정해져 있는 것은, 공부한 분들 외에는 쓰는 것 금지. 문제 선택의 중요성입니다.

인공지능 신뢰성(Trustworthiness)/믿을수있음/AI 신뢰성 검인증 제도(CAT), 한국 추진(TTA), 3단계

토픽 이름 (하)	인공지능 신뢰성(worthiness: 가치)
분류	인공지능 > 평가 > 인공지능 신뢰성(ˈtrʌs(t)ˌwəːðɪnɪs)
키워드(암기)	안전성, 설명가능성, 투명성, 견고성, 공평성, 다양성 / 안설투견공 / 기고체

번호	기출문제/예상문제	회차
1	인공지능 **신뢰성의 개념과 핵심 속성**에 대하여 설명하시오.	**133.관.1.8**

I. 인공지능 활용 및 확산 부작용 방지, 인공지능 **신뢰성 정의**
- AI 시스템이 일관적, 안전하며, 예측 가능한 방식으로 동작하여
사용자·조직·사회가 AI의 판단과 결과를 믿을 수 있는 수준

II. 인공지능 신뢰성의 주요 핵심 속성
(출처: OECD, EU AI Act, ISO/IEC 42001(2023), NIST AI RMF(2023)
가. 인공지능 신뢰성의 핵심 속성 분류도

핵심속성	설명
정확성 (Accuracy·Performance)	- 다양한 환경과 입력에서도 AI가 일관된 정확도를 유지해야 함 - Drift, 오판율, 성능 저하에 대한 지속적 모니터링 필요 - 테스트 데이터 편중을 제거하고 현실 세계(Real-world) 시나리오 기반 검증 필요
안전성(safety)	- AI가 오작동·예상치 못한 행동을 하지 않도록 설계 - 인공지능이 판단·예측한 결과로 시스템이 동작하거나 기능이 수행됐을 때 사람과 환경에 위험을 줄 가능성이 완화 또는 제거된 상태
공정성(Fairness)	- 인공지능이 데이터를 처리하는 과정에서 특정 그룹에 대한 차별이나 편향성을 나타내거나, 차별 및 편향 포함한 결론에 이르지 않는 상태
투명성·설명가능성(Explainability & Transparency)	- 인공지능이 내리는 결정에 대한 이유가 설명 가능하거나 근거가 추적 가능하고, 인공지능의 목적과 한계에 대한 정보가 **적합한 방식으로 사용자에게 전달되는 상태** **- XAI**
보안(Security & Robustness)	- 모델 공격(Adversarial Attack), 프롬프트 인젝션, 데이터 포이즈닝 등에 강해야 함. - 모델 보호(암호화, 워터마킹) 및 안전한 배포 필요. - 입력 조작에도 안정적 행동(Adversarial Robustness) 필요
프라이버시 보호 (Privacy)	프라이버시 보호(Privacy) 학습 데이터에 개인정보가 포함된 경우 보호 필요 Differential Privacy, Federated Learning 등 적용 GDPR, AI Act에서 강하게 요구됨
책임성 (Accountability & Governance)	책임성(Accountability & Governance) AI 오판·실수 발생 시 책임 주체 명확화 개발, 운영, 모니터링 프로세스(ISO 42001 기반) 정립 위험 평가, 영향 평가, 로그 기록, 검증 절차 필요

AI 신뢰성 검인증 제도(CAT), 한국 추진(TTA), 3단계

토픽 이름 (하)	CAT
분류	인공지능 > 평가 > 인공지능 신뢰성(ˈtrʌs(t)ˌwəːðɪnɪs)
키워드(암기)	안전성, 설명가능성, 투명성, 견고성, 공평성, 다양성 / 안설투견공 / 기고체

번호	기출문제/예상문제	회차
1	AI 신뢰성 검인증 제도(CAT)를 설명하시오	137.관.1.8

I. AI 신뢰성 확보를 위한 기술적+윤리적+사회적 인증 제도, CAT(Certification of AI Trustworthiness) 개요

정의	- AI 신뢰성 확보를 위한 사회기술적 요건을 중심으로 기업 내부의 AI 거버넌스 및 주요 신뢰성 특성에 대한 검증, 확인 활동을 1~3 레벨 인증하는 제도

나. 인증대상 – 2가지 암기 /* 인증 대상: '제품·서비스', '조직' 두 가지로 구성 */

인증 대상	적용 표준	시험·심사 내용	인공지능 기본법 관련 대상
제품·서비스	ISO/IEC 23894	인공지능 제품·서비스의 위험관리를 위한 위험관리 프레임워크 및 위험관리 프로세스를 시험·심사	인공지능 시스템
조직	ISO/IEC 42001	인공지능 제품·서비스를 개발하는 조직의 책임있는 AI 개발·운영 역량을 검사하기 위해 인공지능 경영시스템을 심사	인공지능 사업자
조직	ISO/IEC 38507	인공지능 사용 조직이 인공지능 이용(도입) 시, 해당 조직의 거버넌스를 심사	인공지능 사업자, 인공지는 이용자 (개인이 아닌 이용 조직에 한함)

나. AI 신뢰성 인증 제도의 인증 기준 2.0 참조

CAT 2.0

ISO/IEC 23894(제품·서비스)

ISO/IEC 42001(조직)

ISO/IEC 38507(조직)

ISO/IEC TR 24028(임시 공개중)

CAT 1.0

요구사항 1. 인공지능 시스템에 대한 위험관리 계획 및 수행

요구사항 2. 인공지능 거버넌스 체계 구성

요구사항 3. 인공지능 시스템의 신뢰성 테스트 계획 수립

요구사항 4. 인공지능 시스템의 추적가능성 및 변경이력 확보

요구사항 5. 데이터의 활용을 위한 상세 정보 제공

요구사항 6. 데이터 견고성 확보를 위한 이상 데이터 점검

요구사항 7. 수집 및 가공된 학습 데이터의 편향 제거

요구사항 8. 오픈소스 라이브러리의 보안성 및 호환성 점검

요구사항 9. 인공지능 모델의 편향 제거

요구사항 10. 인공지능 모델 공격에 대한 방어 대책 수립

요구사항 11. 인공지능 모델 명세 및 추론 결과에 대한 설명 제공

요구사항 12. 인공지능 시스템 구현 시 발생 가능한 편향 제거

요구사항 13. 인공지능 시스템의 안전 모드 구현 및 문제발생 알림 절차 수립

요구사항 14. 인공지능 시스템의 설명에 대한 사용자의 이해도 제고

요구사항 15. 서비스 제공 범위 및 상호작용 대상에 대한 설명 제공

II. AI 신뢰성 인증 제도의 인증 기준

평가세부 항목	평가 내용
1. AI 시스템에 대한 **위험관리** 계획 및 수행	- **위험관리(Risk Management)** 요소, 위험관리 계획 및 수행 프로세스 명확성 - 위험 식별, 분석, 평가, 대응 과정의 체계적인 수행
2. **AI 거버넌스** 체계 구성	- 거버넌스 체계가 조직 및 AI 모델·시스템의 범위를 명확히 정의하는지 확인 - 거버넌스 체계가 운영에 대한 책임과 권한을 명확히 구분하는지 확인 - 거버넌스 체계가 조직의 제품 및 서비스를 감시, 관리, 감독하는지 확인 - 거버넌스 체계를 지속적으로 평가하고 개선하는지 확인
3. AI 시스템의 **신뢰성** 테스트 계획 수립	- 테스트를 통해 달성하고자 하는 목표의 명확성 - 테스트 대상이 되는 AI 시스템의 기능과 범위의 명확성 - 테스트를 수행하는 데 사용할 방법론(평가기준 포함)이 정의돼 있는지 확인 - 테스트를 수행하는 데 필요한 환경과 데이터가 정의돼 있는지 확인
4. AI 시스템의 추적가능성 및 변경이력 확보	- 데이터 출처 및 변환 과정에 대한 정보가 명확하게 기록돼 있는지 확인 - 모델 학습과정 및 알고리즘 변화에 대한 정보가 기록돼 있는지 확인 - 코드 및 시스템 버전 관리가 체계적으로 이뤄지고 있는지 확인 - 성능 및 결과에 대한 지속적인 추적과 분석이 이뤄지고 있는지 확인 - 변경 사항에 대한 상세 정보가 정확하게 기록돼 있는지 확인 - 변경 사항에 대한 승인 및 검증 프로세스가 명확하게 정의돼 있는지 확인 - 변경 사항에 대한 롤백 및 복구 기능이 정상적으로 작동하는지 확인
5. 데이터의 활용을 위한 상세 정보 제공	- 메타데이터에 중요 관리정보가 모두 포함돼 있는지 확인 - 정제 과정에서의 데이터 특성 변환에 대한 규칙이 명확한지 확인 - 민간정보 또는 개인정보에 해당하거나 일부 포함돼 있는지 확인 - 데이터 출처(데이터의 원본 제공자 또는 수집 방법)가 명확한지 확인

6. 데이터 견고성 확보를 위한 이상 데이터 점검	- 데이터 수집 과정에서 발생하는 오류, 누락, 부정확한 값 등에 대한 내부 검증 활동 확인 - 데이터 최적화 과정에서 발생하는 데이터 변형, 통계적 오류 등에 대한 내부 검증활동 확인 - 데이터 변조 공격 등을 감지하고 방어하기 위한 수단을 구축했는지 확인
7. 수집 및 가공된 학습 데이터의 편향 제거	- 학습 데이터의 편향 특성, 유형, 기준이 명확하게 정의돼 있는지 확인 - 학습 데이터 수집 및 가공 시 편향 제거를 위한 기술 도입의 적정성 확인 - 데이터 라벨링 작업 지침, 교육, 감독 활동의 이행준수 여부 확인
8. AI 오픈소스 라이브러리의 보안성 및 호환성 점검	- 성능, 안정성, 커뮤니티, 문서화 수준 등을 고려한 오픈소스 라이브러리 선정 기준 확인 / - 코드 검사, 취약점 스캔, 침투 테스트 등 보안 취약점에 대한 발견 및 해결 절차 정의 확인 - 라이브러리 버전 확인, API 호환성 검증, 성능 테스트 등 호환성 점검결과 확인
9. AI 모델의 편향 제거	- 모델의 편향 제거 기법 적용을 위한 분석 내용이 식별됐는지 확인 - 편향성 수준에 대한 정량 분석 수행 가능 여부를 확인
10. AI 모델 공격에 대한 방어 대책 수립	- 모델 공격의 영향도 파악 유무 확인 - 가능한 모델 공격 유형 및 공격에 대한 방어 대책수립 여부 확인
11. AI 모델 명세 및 추론 결과에 대한 설명 제공	- 모델 명세가 구체적이고 충분하며 정확한지 확인 - 적용된 설명가능성 기법의 적절성 확인 - 추론 결과에 대한 설명이 이해하기 쉬우며, 신뢰도가 함께 제시되는지 확인
12. AI 시스템 구현 시 발생 가능한 편향 제거	- 데이터 접근 방식 구현 과정 등 소스코드에서 편향 발생 가능성 확인 - 사용자 인터페이스 및 상호작용 방식으로 인한 편향 확인
13. AI 시스템의 안전 모드 구현 및 문제 발생 알림 절차 수립	**- 안전 모드의 작동 기준·조건과 작동 시 시스템 동작 상황을 확인** **- 안전 모드 해제 기준 및 절차 확인** - 문제 감지 절차 및 식별 방법 확인 - 알림 방식 및 내용 확인 - 알림 수신자 및 사람의 개입 시 해당 역할 확인
14. AI 시스템의 설명에 대한 사용자의 이해도 제고	- 사용자 특성에 맞는 설명 방식이 적용됐는지 확인 - 설명 내용이 명확하고 간결하며, 전문용어 사용을 최소화했는지 확인 - 시각 자료, 다양한 설명 방식 등을 활용해 이해도를 높였는지 확인
15. 서비스 제공 범위 및 상호작용 대상에 대한 설명 제공	- 서비스 약관에 면책조항을 포함한 내용이 적절한지 확인 - 이용정책에 허용 가능한 사용과 금지된 사용 관련 내용이 적절한지 확인 - 개인정보보호 정책에 개인정보 수집 및 처리에 관한 내용이 적절한지 확인 - 실제 서비스 이용환경에서 상기 정책들이 올바르게 적용돼 있는지 확인

III. CAT은 성숙도 모델(Maturity Model) 형태로 운영
- CAT 1.0 기준에서는 3단계 수준(Level) 으로 나뉨

CAT 수준	의미
Level 1 – 기본(Basic)	최소한의 신뢰성 요건 충족
Level 2 – 고도화(Advanced)	기술 + 윤리/사회적 요건 반영
Level 3 – 최고(High/Full Trust)	국제 규제와 동등한 수준

CAT 2.0은 ISO/IEC 23894(위험관리), 42001(AI 경영시스템), 38507(AI 거버넌스) 등 국제표준을 바탕으로 설계됐으며 기존의 문서와 절차 중심의 심사에 더해 실제 운영 환경에서의 AI 시스템 대응능력을 평가하는 기 ·성능 시험을 강화했다.

공공부문 초거대 AI 도입·활용 가이드라인 2.0 /25년 4월 개정판

토픽 이름 (하)	공공부문 초거대 AI 도입·활용 가이드라인 2.0
분류	법 지침 가이드 >
키워드(암기)	

번호	기출문제/예상문제	회차
1	공공부문 초거대 AI 도입·활용 가이드라인2.0에 대해서 아래 내용을 설명하시오. 가. 초거대 AI의 **개념과 구성요소** **나. 초거대 AI의 기술요소** 다. 초거대 AI의 도입절차	**137.관.2.4**
2	**최근 초거대 인공지능(AI: Artificial Intelligence) 도입 및 활용에 필요한 사항을 담은 "공공 부문 초거대 AI 도입•활용 가이드라인"이 발표되었다. 다음 항목에 관하여 설명하시오.** **가. 초거대 AI 개념** **나. 초거대 AI 도입 원칙** **다. 초거대 AI 도입 시 사전 고려사항**	**134.컴.2.**

공공부문 초거대 AI 도입·활용 가이드라인 2.0 /25년 4월 개정판

I. 공공부문 초거대 AI 도입·활용 가이드라인 2.0

디지털플랫폼정부 정책방향 연계 추진을 위한, 초거대 AI 도입 및 활용을 위한 가이드라인 목적

1. 초거대 AI 개요

- 1.1 초거대 AI의 개념과 구성 요소
- 1.2 초거대 AI 발전 경과 및 최근 기술 동향
- 1.3 국내 초거대 AI 시장 현황
- 1.4 해외 주요국 AI 정책 동향

2. 공공부문 초거대 AI 추진방향과 활용사례

- 2.1 공공부문 초거대 AI 추진 방향
- 2.1.1 공공AI 3대 전략 목표
- 2.1.2 범정부 초거대 AI 공통기반 구현
- 2.2 공공부문 초거대 AI 활용 사례
- 2.2.1 초거대 AI 적용 서비스 분류
- 2.2.2 서비스 유형별 활용 사례
- 2.2.3 업무 분야별 활용 사례
- 2.2.4 해외 활용 사례

3. 초거대 AI 도입 절차

- 3.1 도입 원칙 및 고려사항
- 3.2 도입 절차
- 3.2.1 데이터 보안 등급
- 3.2.2 클라우드 서비스 구성 방안
- 3.2.3 데이터 학습 방식
- 3.2.4 서비스 도입 방식
- 3.2.5 유지보수 및 운영
- 3.3 초거대 AI 도입 체크리스트

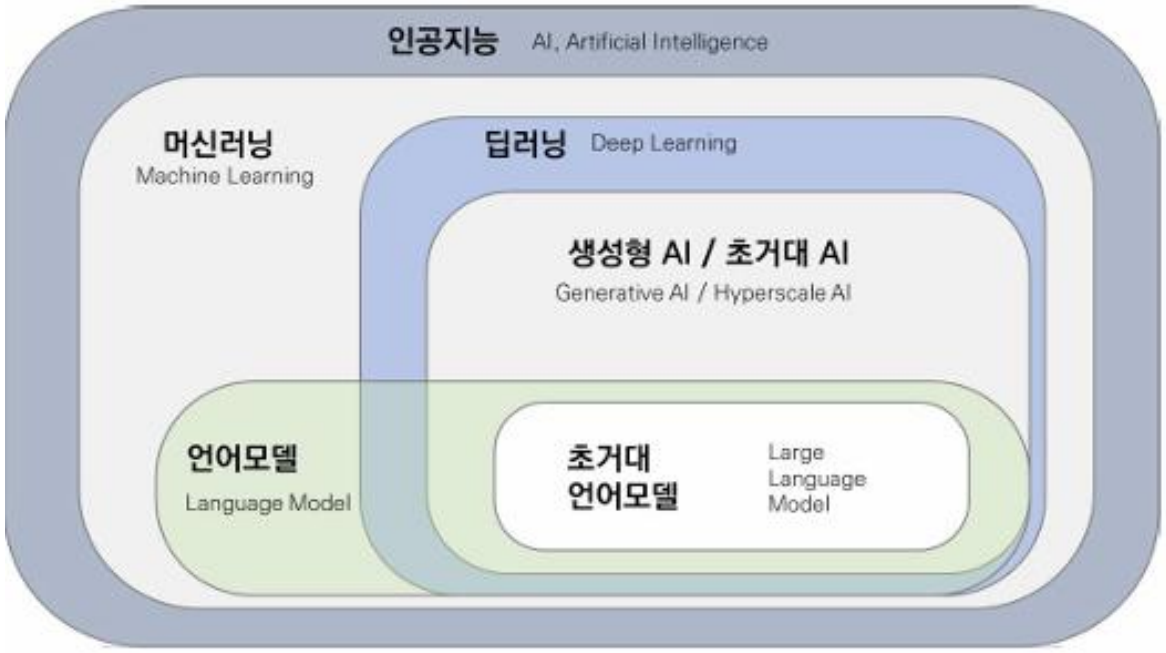
4. 공공부문 AI 성과관리

- 4.1 성과 관리 필요성
- 4.2 AI 성과지표 프레임워크
- 4.3 AI 성과지표 Pool

디지털플랫폼정부 정책방향 연계 추진을 위한, 초거대 AI 도입 및 활용을 위한 가이드라인 목적

구 분	주요 내용(목차)
0. 가이드라인 목적 및 구성	
1. 초거대 AI 개요	1.1. 초거대 AI의 개념과 구성 요소 1.2. 초거대 AI 발전 경과 및 최근 기술 동향 1.3. 국내 초거대 AI 시장 현황 1.4. 해외 주요국 AI 정책 동향
2. 공공부문 초거대 AI 추진방향과 활용사례	2.1. 공공부문 초거대 AI 추진 방향 2.1.1. 공공AI 3대 전략목표 2.1.2. 범정부 초거대 AI 공통기반 구현
	2.2. 공공부문 초거대 AI 활용 사례 2.2.1. 초거대 AI 적용 서비스 분류 2.2.2. 서비스 유형별 활용 사례 2.2.3. 업무 분야별 활용 사례 2.2.4. 해외 활용 사례
3. 초거대 AI 도입 절차	3.1. 도입 원칙 및 고려사항
	3.2. 도입 절차 3.2.1. 데이터 보안 등급 3.2.2. 클라우드 서비스 구성 방안 3.2.3. 데이터 학습 방식 3.2.4. 서비스 도입 방식 3.2.5. 유지보수 및 운영(Operations)
	3.3. 초거대 AI 도입 체크리스트
4. 공공부문 AI 성과 관리	4.1. 성과 관리 필요성 4.2. AI 성과지표 프레임워크 4.3. AI 성과지표 Pool
5. 부록	5.1. 공공 AI 서비스 실증 세부 현황 5.2. 해외 AI 활용사례 인덱스

II. 초거대 AI의 개념과 구성요소
가. 초거대 AI의 개념



(개념) - 초거대 AI로 구현된 언어 모델(LM, Language Model)로 기존 언어 모델보다 훨씬 큰 규모의 모델과 데이터를 활용하여 뛰어난 언어 이해력과 생성 능력을 가진 AI

구성요소	설명
데이터	- AI 학습을 위한 방대한 데이터(텍스트, 이미지 등)
모델 아키텍처	- 주로 트랜스포머 기반의 딥러닝 모델 사용
컴퓨팅 인프라	- GPU, TPU, 슈퍼컴퓨터, 고속 스토리지 등 → 클라우드 컴퓨팅 활용
학습 알고리즘	- 초거대 AI를 학습시키기 위한 알고리즘과 최적화 기법
안정성과 윤리	- AI의 편향 문제 해결, 개인정보 보호, 윤리적 규제 준수

관련용어

용 어	정 의
초거대 AI	대규모 데이터셋을 기반으로 훈련된 딥러닝 모델을 사용하여 문서를 요약, 대조하거나 새로운 콘텐츠를 생성하는 등의 인공지능
초거대 언어 모델 (LLM, Large Language Model)	초거대 규모로 자연어를 학습시킨 인공지능 언어 모델
파운데이션 모델 (Foundation Model)	초거대 AI의 서비스를 제공하기 위해 기초가 되는 모델
파인튜닝 (Fine-tuning)	파운데이션 모델을 특정 작업이나 도메인에 최적화하기 위해 특화된 데이터로 모델의 가중치를 미세하게 조정하여 추가로 학습시키는 방법
사후학습 (Post-training)	파운데이션 모델에 데이터의 최신성과 전문성을 위해 자체 보유데이터로 추가 학습하여, 파운데이션 모델 자체를 고도화하는 방법
검색 증강 생성 (RAG, Retrieval-Augmented Generation)	답변 생성에 있어 외부 리소스를 추가하는 방식의 기술
소규모 초거대 언어 모델 (sLLM, Small Large Language Model)	초거대 언어 모델에 비해 상대적으로 적은 파라미터를 사용하여 학습 시간이나 비용을 절감한 모델
멀티모달 (multimodal)	사람이 기계와 상호 작용할 때 입출력에 텍스트, 음향, 이미지 등 다양한 정보 유형을 통합하여 사용하는 것
AI 에이전트 (AI Agent)	사람이 직접 프롬프트를 입력하는 거대언어모델(LLM) 기반 AI 챗봇과 달리 자율성을 바탕으로 스스로 업무를 수행하는 인공지능
클라우드컴퓨팅서비스	클라우드컴퓨팅을 활용하여 상용(商用)으로 타인에게 정보통신자원을 제공하는 서비스
클라우드컴퓨팅서비스 보안인증 (CSAP, Cloud Security Assurance Program)	「클라우드컴퓨팅 발전 및 이용자 보호에 관한 법률」 제23조의2에 따라 한국인터넷진흥원의 장이 클라우드컴퓨팅서비스의 보안성에 대하여 실시하는 인증
민간 클라우드	「클라우드컴퓨팅법」 제2조제3호에 따른 '클라우드컴퓨팅서비스'로 민간 기업 또는 단체가 제공하는 클라우드컴퓨팅서비스
멀티 클라우드	두 개 이상의 독립적인 클라우드 서비스 제공자가 제공하는 복수의 클라우드컴퓨팅서비스를 연계하여 사용하는 클라우드 배포 모델
디지털서비스 이용지원시스템	「클라우드컴퓨팅법 시행령」 제15조의2제3항에 따라 디지털서비스를 등록 및 관리하는 시스템

IV. 초거대 AI의 도입절차

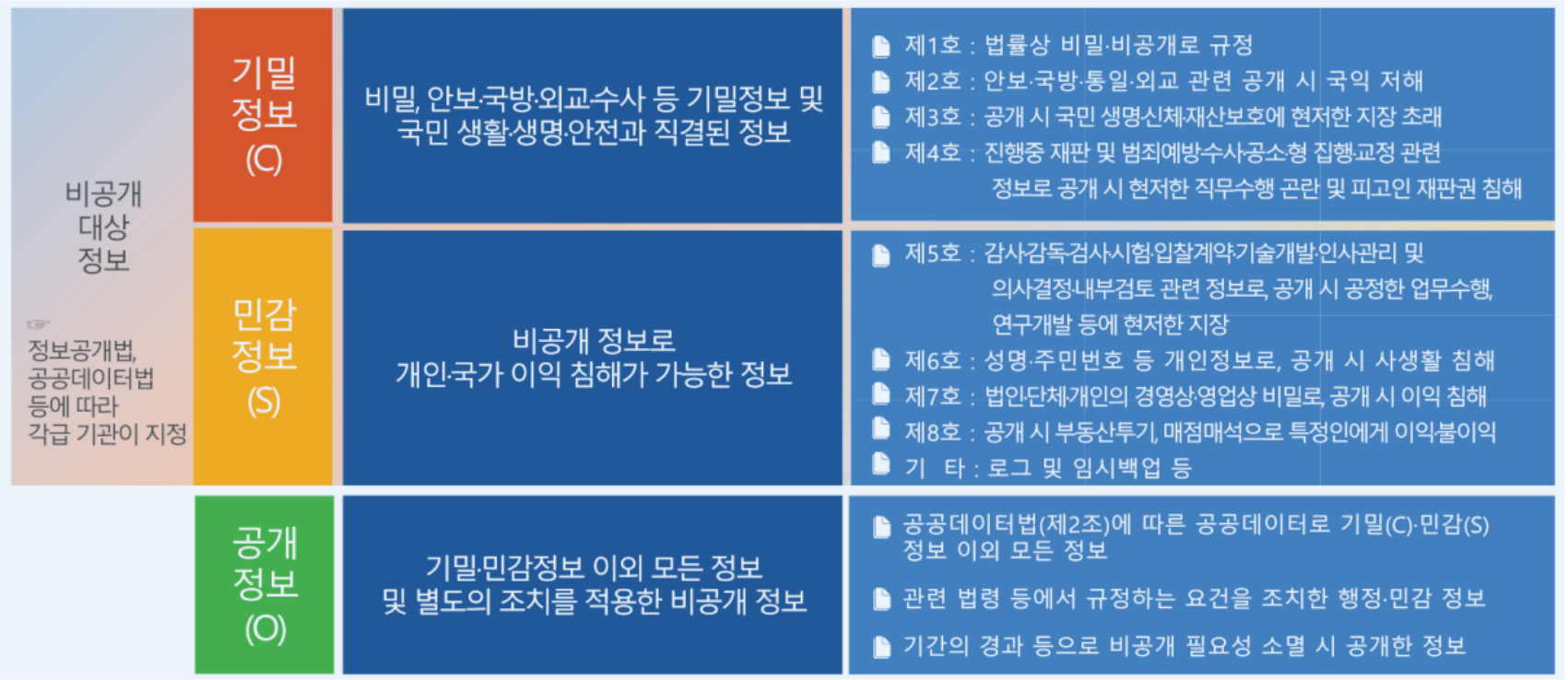
단계	절차	설명
1단계	**데이터 보안 등급**	- 업무 중요도에 따라 **기밀, 민감, 공개 등 3개 등급으로 분류하여** 보안정책 적용

구분	분류 기준	설명
기밀정보 (C)	비밀, 안보·국방·외교수사 등 기밀정보 및 국민 생활·안전과 직결된 정보	- 제1호: 법률상 비밀·비공개로 규정 - 제2호 : 안보·국방·통이일·외교 관련 공개 시 국익 저해 - 제3호: 공개 시 국민 생명·신체·재산·보호에 현저한 지장 초래 - 제4호 : 진행중 재판 및 범죄예방·수사·공소형 집행·교정 관련 정보로 공개 시 현저한 직무수행 곤란 및 피고인 재판권 침해
민간정보 (S)	비공개 정보로 개인·국가 이익 침해가 가능한 정보	- 제5호: 감사·감독·시험·입찰계약·기술개발·인사관리 및 의사결정·매부검토 관련 정보로, 공개 시 공정한 업무수행, 연구개발 등에 현저한 지장 - 제6호: 설명·주민번호 등 개인정보로, 공개 시 사생활 침해 - 제7호 : 법인·단체·개인의 경영상·영업상 비밀로, 공개 시 이익 침해 - 제8호 : 공개 시 부동산투기, 매점매석으로 특정인에게 이익·불이익 - 기타 : 로그 및 임시백업 등
공개정보 (O)	기밀·민감정보 이외 모든 정보 및 별도의 조치를 적용한 비공개 정보	- 공공데이터법(제2조)에 따른 공공데이터로 기밀·민감·정보 이외 모든 정보 - 관련 법령 등에서 규정하는 요건을 조치한 행정·민감 정보 - 기관의 경과 등으로 비공개 필요성 소멸 시 공개한 정보

IV. 초거대 AI의 도입절차

단계	절차	설명
1단계	**데이터 보안 등급**	- 업무 중요도에 따라 **기밀, 민감, 공개 등 3개 등급으로 분류하여** 보안정책 적용

그림 9 업무정보에 대한 C/S/O 분류 기준



* 출처 : 국가정보원 「국가 망 보안체계(National Network Security Framework, N²SF) 보안 가이드라인(Draft)」('25.1월)

2단계	클라우드 구성 방안	- 클라우드 영역 및 규모 선정, 클라우드 도입유형 결정 등

3.2.2 클라우드 서비스 구성 방안

● 민간 클라우드 서비스 구성을 위한 도입 절차

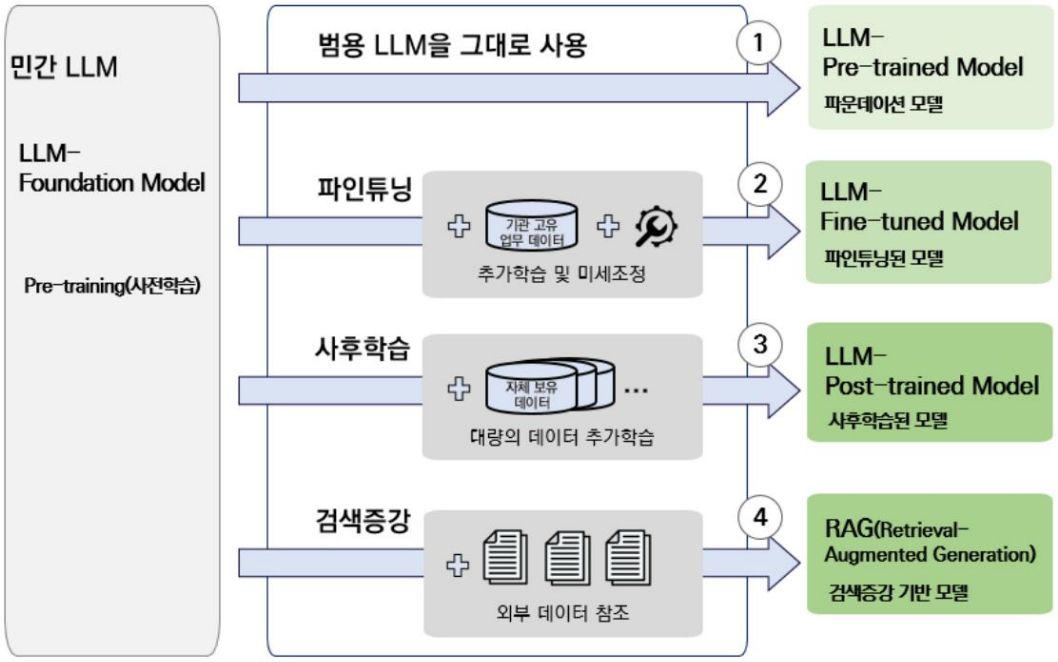
수행 절차	단계별 고려사항
1) 클라우드 영역 및 규모 선정	· 시스템을 운영할 클라우드 영역 선정 · LLM 도입 유형의 인프라 규모를 확인(CPU, GPU, 메모리, 스토리지 등)
2) 클라우드 도입유형 결정	· 클라우드 자원의 자원 공유 여부 확인(기관 단독 사용, 타 기관과 공유) · 클라우드 자원에 대한 직접 물리적 통제 또는 위탁관리 여부 검토
3) 사업 추진 방향 결정	· 보유하고 있는 정보자원과 도입할 정보자원 정의 - 이용 중인 정보자원의 전환, 재구축 - 도입 자원 정의
4) 서비스 벤더 선정	· 보안성, 비용효율성, 안정성, 확장성 검토 후 벤더 선정

3단계	**데이터 학습 방식**	- 파운데이션 모델, 파인튜닝 모델, 사후 학습 모델, RAG 기반 모델로 구분

3.2.3 데이터 학습 방식

● 데이터 학습 방식에 따른 LLM 유형은 크게 파운데이션 모델, 파인튜닝된 모델, 사후학습된 모델, RAG(Retrieval-Augmented Generation, 검색 증강 생성) 기반 모델로 구분할 수 있음

그림 10 학습방식에 따른 LLM 유형



IV. 초거대 AI의 도입절차

4단계	**서비스 도입 방식**	- 디지털 서비스 구매(클라우드 컴퓨팅/융합서비스) / 조달 요청을 통한 용역발주
5단계	유지보수 및 운영(Ops)	- 데이터 준비, 모델 구축, 초기 설정, 사전학습, 추가학습, 교육, 배포, 모니터링, 최적화 등의 운영 관리 및 거버넌스 체계 마련 등
6단계	성과관리	- AI 과제의 체계적인 성과 관리를 위해 성과지표 설정 및 관리

그림 8 도입 절차

3.2.1 **데이터 보안 등급**	업무 중요도에 따라 기밀(Classified), 민감(Sensitive), 공개(Open) 등 3개 등급으로 분류하여 보안정책 적용
3.2.2 **클라우드 구성 방안**	클라우드 영역 및 규모 선정, 클라우드 도입유형 결정 등 클라우드 서비스 구성
3.2.3 **데이터 학습 방식**	파운데이션 모델, 파인튜닝된 모델, 사후 학습된 모델, RAG(검색증강생성) 기반 모델로 구분
3.2.4 **서비스 도입 방식**	디지털 서비스 구매(클라우드 컴퓨팅서비스/융합서비스) 및 조달 용역발주 방식으로 추진
3.2.5 **유지보수 및 운영(Ops)**	데이터 준비, 모델 구축, 초기 설정, 사전학습, 추가학습, 교육, 배포, 모니터링, 최적화 등의 운영 관리 및 거버넌스 체계 마련 등
4 **성과 관리**	AI 과제의 체계적인 성과 관리를 위해 성과지표 설정 및 관리

IV. 초거대 AI의 도입절차

6단계	성과관리	- AI 과제의 체계적인 성과 관리를 위해 성과지표 설정 및 관리

붙임3 공공부문 AI 과제 성과지표 개념도 (예시)

AI 과제 최종목표

[1] 공공서비스 개선
[2] 행정서비스 향상
[3] 사회현안의 해결

=

	투입지표(공통)	과정지표(공통)
	- 전문인력 투입 수 - 인프라 구축 비용 - AI 모델 복잡도 지수 - 데이터 품질 점수 - AI윤리 준수 지수	- 구축률 - 모델성능 - 데이터 품질 개선도 - 알고리즘 최적화 횟수 - 보안 취약점 개선율

AI 서비스	산출지표(고유)	결과지표(고유)
지능형 정보처리	- 정보 분석 정확도 - 데이터 처리량 - 데이터 통합 성공률 - 인사이트 도출 건수 - 정보 요약 품질	- 정보 활용 적시성 개선도 - 분석 결과 활용 비율 증가율 - 데이터 기반 의사결정 비율 증가율 - 부서 간 정보 공유 및 협업 증진도 - 예측 정확도 개선율
업무 자동화	- 의사결정 처리 건수 - 의사결정 소요 시간 - 시스템 안정성 - 판정 일관성 지수 - 의사결정 설명 가능성 점수	- 업무 처리 시간 단축률 - 의사결정 비용 절감률 - 업무 처리 자동화율 - 인적 자원 재배치 효율성 - 데이터 기반 의사결정 비율
대화형 서비스	- 처리된 상담 건수 - 평균 상담 처리 시간 - 질의 응답 정확도 - 상담 만족도 점수 - 첫 응답 해결률	- 상담에 대한 사용자 만족도 - 상담 서비스 접근성 향상률 - 상담 처리 시간 단축률 - 상담 서비스 운영 비용 감소율 - 상황 인식 대응 정확도
모니터링·알람	- 모니터링·알림 건수 - 알림 전달 시스템 응답 시간 - 처리 가능한 데이터 소스 수 - 오탐지율 - 알림 우선순위 정확도	- 알림 정확도 - 이상 징후 조기 발견율 - 사용자 만족도 개선률 - 상황 인식 추천 비율 - 위험 상황 대응 시간 감소율