

공공부문을 위한 지능 기술 및 인프라 분석 프레임워크

발행처 한국지능정보사회진흥원 지능기술인프라본부

발행일 2024. 12. 00

본 자료 복제에는 저작권 소유자의 동의가 필요합니다.
저작권 본 자료의 내용은 한국지능정보사회진흥원(NIA)의 공식
견해와 다를 수 있습니다.

문의 email : ○○○○@nia.or.kr

목차(Contents)

I. 추진 배경	2
II. 프레임워크 개요	5
III. 지능 기술 및 인프라 분석 프레임워크	7
IV. 특화 분석 프레임워크	8
4.1 특화 분석 프레임워크 개요	8
4.2 '생성형 인공지능' 프레임워크	11
4.2.1 생성형 인공지능 프레임워크 세부 요소 정의	
4.2.2 기능 요소	
4.2.3 단계 요소	
4.2.4 품질 요소	

요약(Abstract)

첨단 정보기술의 비약적인 발전은 공공부문 전반에 걸쳐 정보시스템 구축과 디지털 서비스 활용의 복잡성을 심화시키며, 공공서비스 구현 유형을 더욱 다양하고 복잡하게 만들고 있다. 특히 기계학습 알고리즘과 GPU 인프라의 발전으로 초거대 생성형 AI 모델 및 특정 도메인에 최적화된 소규모 AI 모델의 수요도 지속해서 높아지고 있으며, 서비스 제공을 위한 AI 모델 및 인프라 구성의 다양성과 복잡성이 증가하고 있다.

이러한 기술 환경의 복잡성은 공공부문이 최신 기술과 서비스를 효과적으로 활용하여 공공서비스를 제공하는 데 상당한 어려움을 초래하고 있다. 특히 다양한 AI 모델과 알고리즘에 대한 체계적인 분석 정보가 부족하여, AI 기술을 성공적으로 도입하고 활용하는 데 제약이 있다. 공공부문 수요자들은 첨단 AI 기술의 활용을 요구하지만, AI 기술의 빠른 변화 속도를 자체적으로 추적, 관리, 분석하며 활용하는 것은 쉽지 않다.

기하급수적으로 쏟아지는 AI 기술을 분석하고 효과적으로 활용하기 위해서는 정확하고 체계적인 지식 기반이 필요하며, 이러한 지식 기반은 기술의 급격한 발전에 맞춰 지속적으로 업데이트하고 관리되어야 한다. 본 가이드라인은 공공부문에서 지능형 정보시스템을 구축하거나 첨단 서비스를 활용할 때 필요한 기술 분석 및 선정 과정을 체계적으로 지원하는 분석 프레임워크를 제시하는 것을 목표로 한다.

공공부문은 '지능기술 및 인프라 분석 프레임워크'를 활용하여 지능기술을 체계적으로 평가하고 도메인에 적합한 기술을 신속하고 효율적으로 공공서비스에 적용할 수 있다. 즉, 정보화 사업 추진 시 본 프레임워크를 활용 함으로써 설계, 구현, 테스트, 배포, 운영에 이르는 모든 단계에서 지능 기술 및 인프라를 객관적으로 비교, 평가하여 도메인에 최적화된 서비스를 효과적으로 제공할 수 있다.

I. 추진 배경

첨단 정보기술의 비약적 발전과 변화가 공공서비스에 적용되면서, 이를 구현 방식이 점차 복잡화되고 다양화되었으며, 기술적·관리적 측면의 복잡성도 함께 증가하고 있다. 이러한 변화는 공공부문 전반의 정보시스템 구축과 디지털 서비스 활용 방식에 중대한 영향을 미치고 있다.

인공지능(AI) 기술의 급격한 발전은 AI 모델과 알고리즘의 다양성과 복잡성을 심화시키고 있으며, 대규모 AI 모델뿐 아니라 특정 도메인에 최적화된 소규모 AI 모델의 수요도 증가하고 있다. 이러한 AI 모델들은 자연어 처리, 이미지 생성, 컴퓨터 비전 등 다양한 분야에서 활용되고 있다.

대표적인 사례로, 'Hugging Face'에서는 2022년에는 신규 AI 모델 등록이 비교적 느리게 증가했으나, 2023년 이후로 기하급수적으로 증가하여 2024년 11월 기준으로 약 114만 개의 AI 모델을 제공하고 있다. 이는 AI 기술의 발전 속도가 가히 폭발적임을 보여준다.

< Hugging Face LLM 모델의 누적 다운로드 성장¹⁾ >

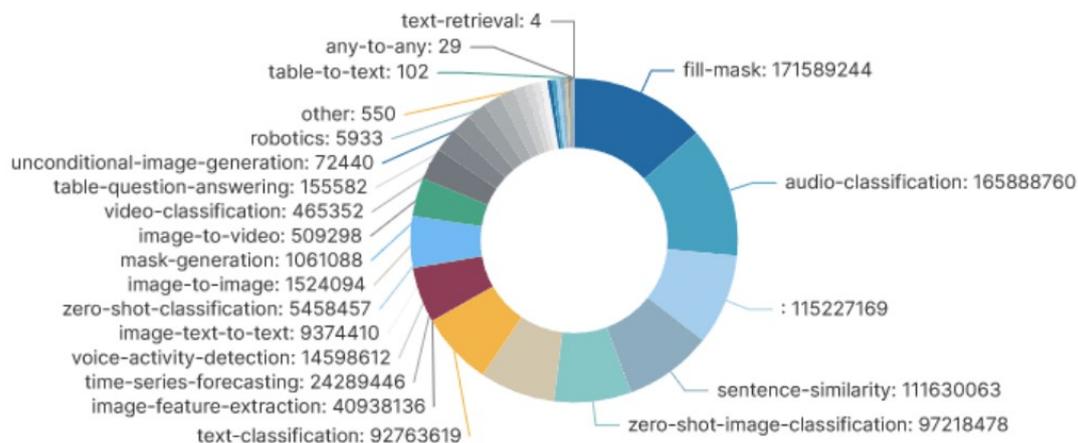


이러한 AI 모델들은 크기(예: 7B, 13B, 80B, 175B, 204B 등), 공개 여부(코드, 가중치, 데이터의 오픈 상태), 기능적 분류(언어, 이미지, 음성 등), 서비스 수준 등 다양한 기준에 따라 구분된다. 이처럼 모델 간의 특성과 분류 방식이 다양해짐에 따라, 전체적인 기술 환경의 복잡성도 점점 증가하고 있다. 아래는 2024년 8월 말부터 한달동안 Hugging Face

1) 자료 : https://gigazine.net/gsc_news/en/20240928-hugging-face-ai-models/

에서 다운로드된 주요 AI 모델의 유형과 사용량을 보여주는 그래프로, 다양한 모델이 실제로 광범위하게 활용되고 있음을 알 수 있다.

< Hugging Face에서 사용된 AI 모델의 주요 유형 및 다운로드 건수²⁾ >



또한 AI 기술을 더욱 효율적으로 응용할 수 있는 도구와 라이브러리의 수요가 증가하고 있으며, AI 모델을 구축하고 운영하는 데 필요한 데이터 처리 도구, GPU 및 고성능 컴퓨팅 인프라, 관련 기술 및 서비스의 활용 구조가 더욱 복잡해지고 있다.

공공부문에서는 첨단 기술들을 도입하여 공공서비스를 개선하고자 하는 강력한 필요성이 있지만, 첨단 정보기술의 비약적인 발전으로 공공부문 전반에 걸쳐 정보시스템 구축과 디지털 서비스 활용의 복합성이 심화되었기 때문에 지능 기술과 인프라를 효율적으로 구성·활용하고 관리하는 데 상당한 난제를 겪고 있다.

공공부문 정보화 담당자들은 확장성, 비용 효율성, 접근성, 업데이트 및 유지보수 등 다양한 기준을 고려하며 클라우드 기반의 AI 서비스 환경구성과 구축형 AI 서비스 환경구성을 검토하고, 구성 방향 따라 세부적인 기술을 분석·도입해야 하는 상황에 직면해 있다. 하지만 다양한 AI 모델, 알고리즘, 데이터 도구, 인프라 등에 대한 체계적인 분석 정보, 즉

2) 자료 : https://gigazine.net/gsc_news/en/20240928-hugging-face-ai-models/

기술 프로파일링이 부족하여 성공적인 기술 도입과 활용이 쉽지 않은 상황이다. 급변하는 지능 기술을 효과적으로 추적하고 관리하며 이를 적절히 활용하기 위해서는 지능 기술에 대한 지속적인 프로파일링 데이터 갱신이 필수적이나, 이러한 작업을 개별 기관에서 자체적으로 수행하기에는 인프라와 자원이 충분하지 않은 것이 현실이다.

공공부문에서의 이러한 문제점은 지능 기술의 체계적이고 종합적인 분석 및 평가를 가능하게 하는 지식 기반의 필요성을 제기한다. 단순히 지능 기술을 도입하는 것에 그치지 않고, 기술의 첨단성, 안전성을 분석하여 각 도메인에 적합한 기술을 도입하고 확장성 있는 인프라 환경을 구성하는 것이 중요하다.

이와 같은 분석을 체계적으로 수행하기 위해서는 지능 기술의 기능, 성능, 품질 등을 철저히 평가하고, 이를 바탕으로 최적의 기술을 선정할 수 있는 체계적인 분석 프레임워크가 필수적이다. 특히, 기술 발전이 급속하게 이루어지는 현 상황에서는 지식 기반을 정기적으로 갱신하고, 기술의 발전에 맞추어 지식을 체계적으로 관리하는 과정이 필요하다. 지능 기술을 도입하기 위해 최신 트렌드를 끊임없이 모니터링·분석하여, 적시에 기술을 활용할 수 있는 체계가 마련되어야 한다.

II. 프레임워크 개요

지능 기술 및 인프라 분석 프레임워크의 목적, 활용 대상, 활용 범위 기대효과는 다음과 같다.

활용 목적

공공부문에서 지능형 정보시스템을 구축하거나 첨단 서비스를 활용할 때 필요한 기술 분석 및 선정 과정을 체계적으로 지원하기 위해 활용한다. 즉, 분석 프레임워크를 활용해 정보화 사업을 발주할 때 지능 기술 및 인프라를 객관적으로 분석·비교하여 서비스 설계부터 구현, 테스트, 배포, 운영에 이르기까지 모든 단계에서 최적의 기술 도입을 돕는다.

활용 대상

본 프레임워크는 공공부문 담당자, 관련 사업 담당자 및 연구자들이 신규 정보화 서비스 기획 또는 개선 시, 자체 구축 또는 클라우드 서비스 기반의 지능 기술 및 인프라 활용을 원하는 경우 활용할 수 있다.

활용 범위

분석 프레임워크는 행정 및 정책, 보건 및 의료, 교육, 사회복지, 교통 및 물류, 환경 및 에너지, 치안 및 국방, 도시 및 지역 개발, 법률 및 사업, 문화 및 관광 등 다양한 공공부문 도메인 응용 분야에 활용될 수 있다.

기대효과

- **공공서비스 품질 향상** : 지능 기술 및 인프라 분석 프레임워크의 도입은 공공서비스의 품질을 전반적으로 향상시킬 수 있다. 다양한 AI 기술과 알고리즘을 체계적으로 분석하고 도메인에 최적화된 기술을 적정 인프라에 적용함으로써 공공서비스의 효율성과 정밀성을 크게

개선할 수 있다. 이는 공공서비스 제공 과정에서의 신속성과 정확성을 증대시키는 결과를 가져오며, 공공부문 수요자의 요구를 보다 효과적으로 충족시키는 맞춤형 서비스 제공을 가능하게 한다.

- **기술 도입 및 활용의 효율성 증대** : 첨단 지능 기술의 복잡성을 체계적으로 관리할 수 있는 분석 프레임워크는 공공부문이 AI 기술과 서비스를 신속히 도입하고 활용할 수 있는 기반을 제공한다. 정보화 사업 전 과정에서 이 프레임워크를 활용하면 설계부터 운영까지 기술과 인프라를 객관적으로 평가하고 비교할 수 있다. 이를 통해 공공부문은 최적의 기술을 효과적으로 선정할 수 있으며, 급변하는 기술 환경에서도 유연하고 효율적인 대응이 가능하다.
- **공공부문의 기술 역량 강화** : AI 기술 및 인프라에 대한 체계적인 평가와 지식 기반을 제공함으로써 공공부문의 기술 의사결정 역량이 강화된다. 분석 및 선정 과정의 표준화를 통해 기술 활용의 신뢰성과 전문성이 제고되며, 이는 공공부문이 첨단 기술 도입 시 발생할 수 있는 시행착오를 줄이는 데 기여한다. 또한, 최신 기술 트렌드와 서비스 변화에 대한 지속적인 업데이트를 통해 공공부문 내 기술 활용의 경쟁력이 한층 강화될 것으로 기대된다.

III. 지능 기술 및 인프라 분석 프레임워크

‘분석 프레임워크’는 공공부문에서 발주자가 지능정보 시스템을 구축하거나 첨단 서비스를 활용할 때 다양한 지능 기술을 최적으로 선정할 수 있도록 지원하는 도구이다. 본 프레임워크는 크게 ‘기능’, ‘단계’, ‘품질’로 구성되어 있다. 각 구성요소는 특성에 따라 세부 요소들로 구체화된다.

분석 프레임워크의 주요 요소의 정의는 다음과 같이 정의한다.

- **기능 요소**는 지능 기술 및 서비스를 제공하는 고유한 역할과 응용을 말한다.
- **단계 요소**는 지능 기술 및 서비스가 구현되는 단계를 말하며, 크게 준비, 개발, 배포, 제공, 운영 등으로 나누어진다.
- **품질 요소**는 지능 기술, 서비스, 인프라의 성능(기능·도메인별)이나 비용 등과 같은 비기능적 품질(정량적, 정성적)을 의미한다.

< 특화 분석 프레임워크 구성 >

주요 요소	분류	세부 요소
기능 요소 (Function Factor)	분류 1	기능 관련 세부요소 1 (별도 정의)
	분류 2	기능 관련 세부요소 2 (별도 정의)
 (확장 가능)
단계 요소 (Process Factor)	분류 1	단계 관련 세부요소 1 (별도 정의)
	분류 2	단계 관련 세부요소 2 (별도 정의)
 (확장 가능)
품질 요소 (Quality Factor)	분류 1	품질 관련 세부요소 1 (별도 정의)
	분류 2	품질 관련 세부요소 2 (별도 정의)
 (확장 가능)

IV. 특화 분석 프레임워크

4.1 특화 분석 프레임워크 개요

지능 기술은 응용 범위가 매우 넓고, 각각 고유한 기능, 구현 단계, 품질 요소를 가진다. 이러한 특성으로 포괄적인 분석 도구만으로는 모든 지능 기술에 대한 구체적인 분석이 어려울 수 있다. 일례로 판별을 위한 지능 기술과 생성을 위한 지능 기술은 세부적인 품질이나 기능 요소가 서로 상이할 수 있다.

특화 분석 프레임워크는 기존 분석 프레임워크를 기반으로 특정한 지능 기술 분야에 적합한 세부 요소들을 정의한 것이다. 이를 통해 특정 기술 분야에 적합한 맞춤형 분석과 평가가 가능하다.

특화 분석 프레임워크는 지능 기술의 기능적 특성을 명확히 구분하고 설명한다. 예를 들어, 생성형 인공지능과 전통적 머신러닝은 서로 다른 기능을 수행하므로, 해당 기술 특화된 프레임워크는 각 기술별 특성에 부합하는 분석을 지원한다.

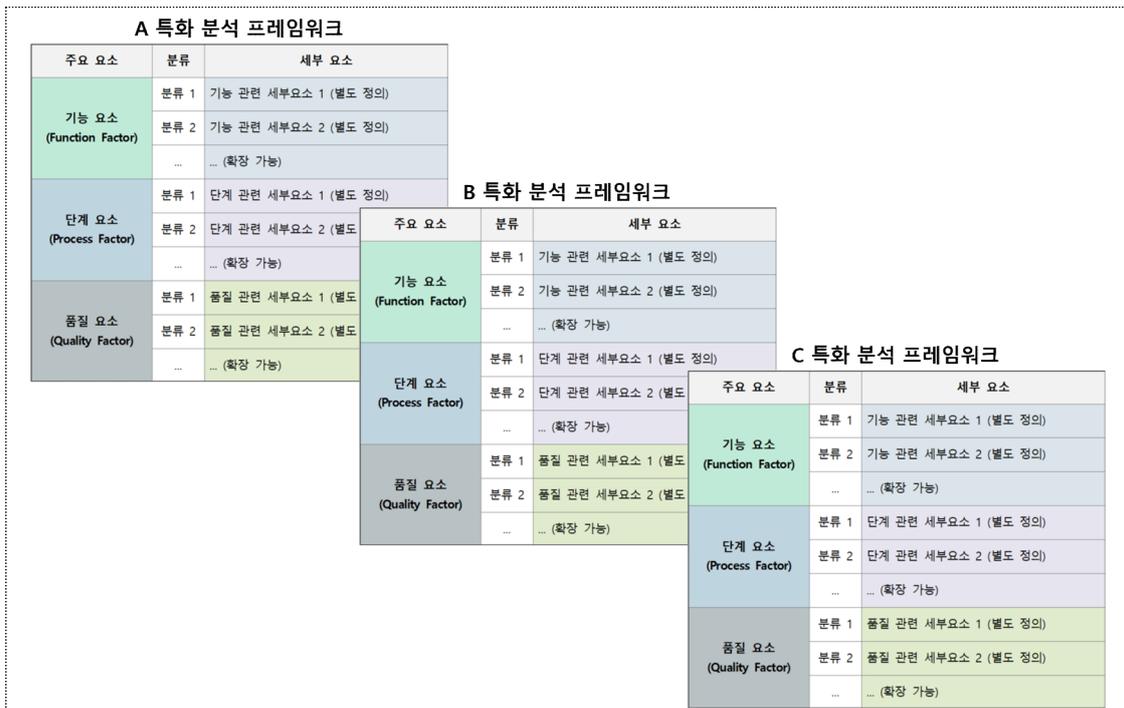
또한, 기술 구현 단계의 복잡성을 체계적으로 이해하고 관리할 수 있도록 돕는다. 지능 기술은 여러 단계(준비, 개발, 배포, 운영 등)로 구성되며, 각 단계는 독립적이지만 상호 연관성을 가진다. 특화 분석 프레임워크는 단계별 요소를 구체화하여 기술이 어느 단계에서 어떻게 활용되는지 명확히 분석할 기준을 제공한다.

마지막으로, 품질 요소를 심층적으로 분석할 수 있다. 지능 기술의 성능은 정확도나 처리 속도뿐만 아니라 신뢰성, 확장성, 적응성 등의 요소도 고려해야 한다. 특화 분석 프레임워크는 특정 지능 기술 분야별 품질 요소를 정의하고 세부 측정 기준을 제공함으로써 기술의 전반적인 품질을 다각도로 평가하거나 이해할 수 있도록 한다.

결과적으로, 특화 분석 프레임워크는 특정 지능 기술 분야를 구성하는 다양한 실제 기술들을 그 분야에 적합한 기능, 단계, 품질 관점에서 맞춤형으로

분석하여 체계적으로 이해할 수 있는 틀을 제공함으로써 정밀한 평가와 기술 선정을 가능하게 한다.

< 특화 분석 프레임워크 구성 >



본 프레임워크에서는 중점 분석이 필요한 특화 분야를 생성형 인공지능 분야, 머신러닝 분야, 데이터 도구 분야, 지능 인프라 분야로 분류한다. 또한, 본 프레임워크는 생성형 인공지능에 대한 특화 분석 프레임워크를 우선적으로 정의하며, 이후 개정을 통하여 타 특화 분야로 대상을 점진적으로 확대해 나갈 계획이다.

- 1) **생성형 인공지능** : 주어진 입력 데이터나 프롬프트를 기반으로 새로운 콘텐츠를 생성하는 인공지능 기술을 의미한다. 이 기술은 텍스트, 이미지, 음악, 코드 등 다양한 형식의 데이터를 생성할 수 있다.
- 2) **머신러닝** : 컴퓨터가 데이터를 통해 학습하고 예측하거나 결정을 내릴 수 있는 기술로, 데이터 분석을 통해 데이터에서 패턴을 찾아내어 새로운 데이터에 대한 예측이나 작업 수행을 가능하게 한다.

- 3) **데이터 도구** : 데이터 수집, 저장, 관리, 분석, 시각화, 처리를 지원하는 소프트웨어나 플랫폼으로, 의사결정을 돕고 통찰력을 도출하며 다양한 작업의 자동화에 활용된다.
- 4) **지능 인프라** : AI 모델의 개발, 학습, 배포, 운영을 효율적으로 지원하기 위해 최적화된 디지털 자원의 집합이다. 컴퓨팅 성능, 데이터 저장소, 네트워크 연결성, 소프트웨어 플랫폼 등을 포함하며, 기술의 확장성과 효율성을 높이는 데 핵심적인 역할을 한다.

< 분석 프레임워크 구성 >



※ 특화 프레임워크는 기능, 단계, 품질 요소를 기준으로 분석 대상의 특성에 맞게 세분화하며, 세부 요소로 확장 구성한 프레임워크를 단계적으로 확대 예정

4.2 ‘생성형 인공지능’ 프레임워크

생성형 인공지능은 수십억 개에서 수천억 개 이상의 매개변수를 가진 인공지능 모델로, 대규모 데이터를 학습하여 자연어 처리, 이미지 인식, 텍스트 생성, 번역 등 다양한 작업을 수행한다. 대표적으로 OpenAI의 GPT 시리즈, Meta의 LLaMA 등이 있다.

이 특화 분석 프레임워크는 분석 프레임워크의 생성형 인공지능의 특성에 맞춰 기능, 단계, 품질을 기반으로 세부 요소를 확장한다. 이를 통해 초거대 생성형 AI 모델을 체계적으로 분석하고 평가하는 데 활용할 수 있다. 본 장에서는 생성형 인공지능 프레임워크로 지칭한다.

4.2.1 생성형 인공지능 프레임워크 세부 요소 정의

생성형 인공지능 프레임워크는 생성형 인공지능 모델의 주요 특성을 분석하기 위해 인공지능 모델에 대한 기본 정보와 기능, 단계, 품질 요소에 대한 분석 기준을 제공한다. 인공지능 모델에 대한 기본 정보는 아래와 같은 항목들을 포함한다.

- **이름(Name):** AI 모델이나 기술의 고유 명칭을 지칭한다. 예를 들어 GPT-4, BERT, Stable Diffusion 등이 있다.
- **파라미터 수(Parameters):** AI 모델이 학습할 때 사용하는 조정 가능한 변수의 총합을 의미한다. 파라미터 수가 많을수록 더 복잡하고 성능이 우수하다.
- **조직(Organization):** AI 모델을 개발하거나 배포한 기관 또는 회사를 의미한다. 예시로 OpenAI, Google, Meta 등이 있다.
- **국적(Nationality):** 모델을 개발한 조직의 국가를 나타낸다. 예를 들어 미국, 대한민국, 영국 등으로 표시된다.
- **모델종류(Category):** 모델이 오픈소스로 제공되어 다운로드·설치가 가능한 모델인지 혹은 소스코드가 비공개인 폐쇄형으로 서비스 형

태로만 이용가능한 모델인지를 구분한다.

- **배포일시(Created Date):** AI 모델이 처음 공개되거나 배포된 날짜로, 모델의 최신성이나 발전 상태를 평가하는 데 유용하다.

실질적 분석을 위한 구성 요소인 기능, 단계, 품질은 각각 다음과 같이 세분화된다. 기능 요소는 분야, 모달리티, 지원 언어, 모델 유형, 제공 방식으로 세분화된다. 단계 요소는 생성형 인공지능의 부문 특성에 따라 학습 단계로 국한된다. 마지막으로 품질 요소는 성능, 비용, 개방성의 요소로 세분화된다.

전체 구성 요소는 아래 표와 같다.

< 생성형 AI 모델 분석을 위한 특화 분석 프레임워크 >

주요 요소	세부 요소 분류	세부 요소	
-	개요	· 이름(Name)	· 파라미터 수(Parameters)
		· 조직(Organization)	· 국적(Nationality)
		· 카테고리(오픈소스 모델/클로즈드 서비스 모델 분류)	
		· 배포일시(created date)	
① 기능	분야	· 모델 유형(model type)	· 작업(task)
	모달리티	· 입력(input)	· 출력(output) · 모달리티 종류
	언어	· 한국어 지원	· 지원 언어(support lang)
	모델 유형	· 언어 모델 유형	· 부모 모델
	제공 방식	· 모델 제공 방식(Deployment channel)	
② 단계	모델 빌드&학습	· AI 모델 단계	
③ 품질	성능	· 파라미터 수	· 한국어 성능
		· 분야별 지식 수준	· 종합 전문지식 수준
		· 휴먼 선호도	· 추론력
		· 수학 추론 성능	· 코딩 역량
		· AI 신뢰성	· AI 처리 속도
	비용	· 학습 하드웨어(GPU)	· 학습 시간 (training time)
		· 권장 추론 하드웨어(GPU)	· 예상 클라우드 이용 비용
		· 입력 1M 토큰당 비용	· 출력 1M 토큰당 비용
		· 예상 GPU 소비 전력	
	개방성	· 사전학습 데이터	· 사전학습 데이터 상세정보
		· 파인튜닝 데이터	· 파인튜닝 데이터 상세정보
		· 모델 가중치 (LLM weights)	
		· 모델 아키텍처 (Model architecture)	
		· 코드 공개(Code open)	
· 라이선스(License)			

4.2.2 기능 요소

- 1) **분야** : AI 모델이 적용되는 특정 산업이나 분야를 의미하며, 예를 들어 의료, 금융, 교육 등 다양한 분야에서 활용될 수 있다.
 - **모델 유형(Model Type)**: AI 모델의 구조나 설계 방식에 따라 분류된 유형을 지칭한다. 언어 모델, 이미지 분류 모델 등 다양한 유형이 존재하며, 각 모델은 특정 작업 수행을 위해 설계된다.
 - **작업(Task)**: AI 모델이 수행하는 특정 작업을 의미한다. 언어 모델의 경우 문장 완성, 번역, 감정 분석 같은 작업이 포함될 수 있으며, 작업의 종류에 따라 모델의 목적과 성능 평가 기준이 달라진다.
- 2) **모달리티** : AI 모델이 처리하는 텍스트, 이미지, 오디오 등 데이터 유형을 의미한다.
 - **입력(Input)**: AI 모델이 처리하는 데이터의 유형을 의미한다. 텍스트, 이미지, 오디오, 비디오 등이 포함된다.
 - **출력(Output)**: AI 모델이 생성하거나 예측하는 데이터 유형을 의미한다. 출력은 입력 데이터에 따라 다르며, 텍스트 응답, 이미지 생성, 숫자 예측 등 다양한 형태로 나타날 수 있다.
 - **모달리티 종류(Modality type)** : 개별 모달리티(Unimodal), 멀티 모달리티(Multimodality), 교차 모달리티(Cross-Modality), 바인딩 모달리티(Binding Modality) 등 모달리티의 유형을 의미한다.
- 3) **언어** : AI 모델이 한국어를 이해하고 처리할 수 있는지 여부와 모델이 지원하는 언어 목록을 포함한다.
 - **한국어 지원(Korean Supported)**: AI 모델이 한국어를 이해하고 처리할 수 있는지를 나타낸다. 한국어 지원 모델은 번역, 질의응답, 요약 등 한국어 기반 작업을 수행할 수 있다.
 - **지원 언어(Supported Languages)**: AI 모델이 지원하는 언어 목록

을 의미하며, 모델이 여러 언어로 작업할 수 있는지를 보여준다. 다국어 모델은 다양한 언어 사용자에게 유용하다.

4) **모델 유형** : AI 모델의 구조나 설계 방식을 기준으로 분류된다.

- **언어 모델 유형(Language Model Type)**: ① 방대한 텍스트 데이터를 학습하여 자연어 처리 작업을 수행하는 LLM(Large Language Model)과 ② 텍스트, 이미지, 소리, 영상 등 여러 데이터를 동시에 처리하는 LMM(Large Multimodal Model)로 구분된다.
- **부모 모델(Parent Model)** : 여러 하위 모델(자식 모델)을 생성하거나 훈련하기 위해 사용되는 기반 모델을 의미한다.

5) **제공 방식** : AI 모델은 다양한 방식으로 제공될 수 있으며, 이는 사용자의 요구와 환경에 따라 선택된다.

- **모델 제공 방식(Deployment Channel)**: AI 모델이 제공되는 방식에 나타낸다. API 서비스, 온프레미스 설치, 클라우드 기반 서비스 등이 있으며, 사용자 편의성과 요구사항에 따라 선택된다.

4.2.3 단계 요소

1) **모델 빌드 & 학습** : AI 모델을 구축하고 학습시키는 단계로, 모델이 원하는 작업을 수행할 수 있도록 데이터와 알고리즘을 통해 훈련되는 과정이다.

- **AI 모델 단계** : AI 모델은 크게 두 가지 주요 단계로 구분된다.
 - **파운데이션 AI 모델(Foundation AI Model)** : 대규모 데이터로 사전 학습된 기본 모델로, 다양한 작업을 수행할 수 있는 범용적인 기능을 갖춘 모델이다. 이 모델은 특정 작업을 위해 추가적인 학습이나 미세 조정이 가능하다.
 - **Instruction-tuned 모델** : 사용자의 지시나 요청을 명확히 이해하고 수행하도록 특수하게 조정된 모델로, 주로 자연어 처리 분야에서 사용된다. 이 모델은 인간의 지시를 해석하고, 요구에 맞게

응답하는 방식으로 설계된다.

4.2.4 품질 요소

- 1) 성능 : AI 모델의 성능을 측정하는 주요 지표로, 다양한 분야에서 모델의 우수성과 효율성을 평가한다.
 - **파라미터 수(Parameters)**: AI 모델이 학습한 매개변수의 총 수로, 모델의 복잡도와 성능을 나타내는 중요한 지표이다. 모델의 크기나 성능을 평가할 때 중요한 요소로 작용한다.
 - **한국어 성능** : 거대 언어모델의 한국어 성능을 평가하는 벤치마크로, 주요 평가지표는 다음과 같다.
 - **KMMLU** : 한국어에 특화된 다분야 학습 평가 벤치마크로, 다양한 분야에서 한국어 기반 지식을 활용한 논리적 추론 및 학습 능력을 테스트하며, 한국어 자연어 이해 성능을 측정하는 데 사용된다.
 - **Horangi Leaderboard** : 거대 언어모델(LLM)의 한국어 능력을 평가하기 위한 도구로, Q&A 형식의 언어 이해와 대화를 통해 생성 능력을 평가하는 MT-Bench 측정 도구를 제공한다.
 - **분야별 지식 수준** : AI 모델이 다양한 분야에서 지식을 얼마나 잘 이해하고 처리하는지를 평가하는 지표이다.
 - **MLU-Pro** : 여러 언어에서 다분야 학습 능력을 평가하는 리더보드로, 자연어 처리 모델이 역사, 과학, 수학 등 다양한 주제를 얼마나 잘 이해하고 응답하는지를 평가한다. 이를 통해 모델의 다국어 이해 및 학습 능력을 검증할 수 있다.
 - **MMMU** : 다중 모달 학습과 이해를 평가하는 벤치마크로, 텍스트, 이미지 등 다양한 형태의 데이터를 함께 처리하는 AI 모델의 성능을 측정한다. 여러 입력 모달을 통합하여 정확하게 예측하고 이해하는 능력을 평가한다.
 - **종합 전문지식 수준** : AI 모델이 복잡한 질문에 대해 정확하고 일관된 답변을 제공하는 능력을 평가하는 지표로, 모델이 질문의 문맥을

이해하고 다양한 정보를 기반으로 논리적이고 사실적인 답변을 생성하는지, 그리고 복잡한 추론을 처리하는 능력을 측정한다.

- **휴먼 선호도** : AI 모델의 성능을 평가할 때, 인간의 선호도를 반영하여 모델 간의 상대적인 우수성을 측정하는 지표이다.
 - **아레나 점수(Arena Score)** : 챗봇 아레나에서 모델의 성능을 평가하는 점수로, 사용자들이 두 익명화된 모델이 동일한 질문에 대해 생성한 답변을 비교하고 선호하는 답변에 투표하여 모델의 우수성을 평가한다.
 - **투표(Votes)**: 챗봇 아레나에서 사용자가 두 모델의 답변을 비교해 선호하는 답변에 투표하는 방식으로, 이를 통해 모델 간 성능을 객관적으로 비교하고 크라우드소싱을 통해 실제 사용 환경에서의 성능을 평가한다.
- **추론력(Reasoning)** : 모델이 주어진 정보에 따라 논리적인 결론을 도출하고, 멀티스텝 추론을 통해 복잡한 문제를 해결하는 능력이다. 이 능력은 모델이 다양한 상황을 분석하고, 다단계 사고 과정을 통해 정확한 결론에 도달할 수 있는지 평가하는 데 중요하다.
- **다국어 추론(MGSM)** : 여러 언어에서 모델의 논리적 추론 능력을 평가하는 벤치마크로, 다양한 언어로 주어진 질문에 일관된 추론 수행을 확인하여, 언어적 편향 없이 추론 능력 유지에 대해 검증한다.
- **수학 추론 성능** : LLM의 수학적 성능을 평가하는 지표로, 초등 수준부터 고급 수학(경시대회 수준) 문제까지 다양한 난이도의 평가를 포함한다.
 - **GSM-8k** : 수학 문제를 해결하는 모델의 논리적 추론 능력을 테스트하는 데이터셋으로, 초등학교와 중학교 수준의 문제들로 구성되어 있으며 단계적 접근과 논리적 사고 능력이 요구된다.
 - **MATH** : 대수, 기하, 확률 등 다양한 수학 영역의 복잡한 문제를 해결하는 모델의 성능을 평가하는 벤치마크로, 고급 수학적 이해와 정확한 계산 능력을 시험한다.

- **코딩역량** : AI 모델이 코딩 문제를 해결하는 능력을 평가하는 벤치마크로, 프로그래밍 언어와 관련된 다양한 작업에서 모델의 정확성과 효율성을 검증한다.
 - **HumanEval** : 주로 Python을 기반으로 코딩 문제를 해결하는 모델의 성능을 측정하는 벤치마크로, 작성된 코드의 정확성과 효율성을 평가한다.
 - **Big Code Bench** : 대규모 코드 생성 및 분석 능력을 평가하는 벤치마크로, 코드 완성, 오류 감지 등 다양한 프로그래밍 언어 관련 작업에서 AI 모델의 성능을 검증한다.
 - **AI 신뢰성** : AI 시스템이 특정 조건에서 안전하고 일관되며, 예측 가능한 방식으로 작동하는지 평가하는 과정으로, 편향성, 안전성, 및 윤리적 책임을 포함한 다양한 요소를 고려한다.
 - **환각대응추론(TruthfulQA)** : AI 모델이 사실적이고 진실된 답변을 생성하는 능력을 평가하는 벤치마크로, 모델이 환각 현상(허구적이거나 잘못된 정보를 생성하는 현상)에 대한 대응력을 시험하며, 정직하고 정확한 답변을 제공하는지를 검증한다.
 - **Enkrypt Safety Score** : AI 모델이 생성하는 출력의 안전성과 윤리적 적합성을 평가하는 지표로, 유해하거나 위험한 콘텐츠 없이 사회적으로 안전한 출력을 제공하는지 능력을 측정한다.
 - **Enkrypt Risk Score** : AI 시스템의 출력이 초래할 수 있는 잠재적 위험성을 평가하는 지표로, 모델의 신뢰성을 높이는 데 필요한 위험 요인을 분석하고 개선점을 도출한다.
 - **AI 처리 속도** : AI 모델이 1초당 생성할 수 있는 출력 토큰(Output Tokens per Second)의 수로 측정된다. 이 지표는 모델의 처리 성능을 나타내며, 특히 응답 시간이 중요한 응용 프로그램에서 필수적인 성능 요소로 평가된다.
- 2) **비용** : AI 모델을 구축하고 운영하는 데 드는 전체 비용을 평가하는 주요 지표로, 개발부터 배포, 유지까지의 경제적 부담을 측정한다.

- **학습 하드웨어(GPU)(Training Hardware)** : AI 모델 학습에 사용되는 GPU나 기타 하드웨어를 지칭하며, 대규모 데이터와 복잡한 모델을 빠르게 학습하도록 지원한다.
 - **학습 시간(Training Time)** : AI 모델 학습에 소요되는 총 시간으로, 데이터 크기, 모델 복잡성, 하드웨어 성능 등에 따라 달라진다.
 - **권장 추론 하드웨어(GPU)(Recommended Inference Hardware)** : 학습된 AI 모델이 실시간 응답 생성이나 예측 수행 시 사용할 가장 적합한 하드웨어를 의미한다.
 - **예상 클라우드 이용 비용(Expecteded cloud usage costs)** : 클라우드 플랫폼에서 AI 모델 훈련 및 서비스를 제공할 때 발생하는 비용으로, GPU 사용, 데이터 스토리지, 네트워크 트래픽 등을 포함한다.
 - **입력 1M 토큰당 비용(Input costs per 1M Tokens)**: AI 모델이 100만 개의 입력 토큰을 처리하는 데 드는 비용으로, AI 모델의 경제 효율성을 평가하는 지표다.
 - **출력 1M 토큰당 비용(Output costs per 1M Tokens)**: AI 모델이 100만 개의 출력 토큰을 생성하는 데 드는 비용으로, 운영 효율성과 관련이 있으며, 기존 비용 대비 성능 분석 시 고려된다.
 - **예상 GPU 소비 전력(expected GPU power consumption)** : GPU가 학습 및 추론 작업 중 사용하는 전력량을 의미하며, 에너지 효율성을 판단하는 데 사용된다.
- 3) **개방성** : AI 모델의 개방성은 모델의 학습 과정과 사용 가능한 데이터, 공개된 자료에 대한 접근성을 평가하는 지표로, 공공부문에서 모델을 재사용하거나 개선할 때 고려해야 할 요소들을 포함한다.
- **사전학습 데이터(Pre-training Data)**: AI 모델이 처음 학습할 때 사용하는 대규모 데이터셋으로, 모델이 언어나 기타 입력 데이터의 일반적인 패턴을 학습하는 데 사용된다. 본 세부 요소는 사전학습 데이터의 공개 여부를 의미한다.

- **사전학습 데이터 상세정보(Pre-training data properties)** : 사전학습 데이터의 속성에 따라 모델의 기능을 예측할 수 있으며, 공개된 데이터의 특성(멀티턴(Multi-turn) 가능 여부, 모델 학습에 활용된 데이터 정보 등)을 포함한다.
- **파인튜닝 데이터(Fine-tuning Data)** : 모델의 성능을 특정 작업에 맞춰 개선하기 위해 사용하는 데이터다. 사전 학습 후에 추가 학습을 통해 모델이 구체적이고 맞춤형 작업을 수행할 수 있도록 조정된다.
- **파인튜닝 데이터 상세정보(Fine-tuning data properties)** : 파인튜닝 데이터의 속성에 따라 모델의 기능을 예측할 수 있으며, 공개된 데이터의 특성(멀티턴(Multi-turn) 가능 여부, SFT 단계의 데이터셋 정보, 모델 학습에 활용된 데이터 정보 등)을 포함한다.
- **모델 가중치(LLM Weights)** : 대규모 언어 모델(LLM)의 학습된 매개변수를 의미하며, 모델이 입력 데이터를 바탕으로 예측을 생성하는 데 중요한 역할을 한다. 모델의 성능과 품질에 큰 영향을 미치며, 공개된 모델의 경우 가중치 파일이 제공되어 모델을 재사용하거나 도메인에 맞게 추가 학습을 통해 맞춤형 서비스를 개발할 수 있다.
- **모델 아키텍처(Model Architecture)** : AI 모델의 구조와 설계 방식을 지칭하며, 모델이 데이터를 처리하고 학습하는 방법을 정의한다.
- **코드 공개(Model code)** : 모델을 학습할 때 활용한 코드 및 추론 코드의 공개 여부를 의미한다.
- **라이선스(License)** : AI 모델이나 소프트웨어를 사용할 때 적용되는 법적 조건을 의미한다.