AI 파운데이션 모델 중심 인공지능 연구 동향의 변화

Emerging Trends in AI Research Driven by Foundation Models

황중원 (J.-W. Hwang, jwhwang@etri.re.kr) 시각지능연구실 선임연구원
윤기민 (K.M. Yun, kimin.yun@etri.re.kr) 시각지능연구실 선임연구원
한동현 (D.H. Han, mpolio2@etri.re.kr) 시각지능연구실 석사후연수연구원
배유석 (Y.S. Bae, baeys@etri.re.kr) 시각지능연구실 책임연구원

ABSTRACT

Recent advances in artificial intelligence (AI) have revealed a shift in research practices. Researchers are now constructing shared foundation models with strong generalization and emergent capabilities instead of independently developing domain-specific models. This study surveys the transformations driven by this shift, outlines their conceptual backgrounds, introduces two main approaches to leverage them, states the key structural limitations and proposed remedies, and presents the resulting changes in training, data, and evaluation paradigms. Rather than listing the advances, we emphasize the motivations and implications behind them, offering perspectives on how foundation models reshape the research ecosystem and inform future directions in AI.

KEYWORDS Agentic AI, Benchmark, CoT, Datacentric AI, DPO, Foundation Model, ICL, Instruction Tuning, LLM, LoRA, Multimodal LLM, RAG, RLHF, RLAIF, VLA, World Model

I. 서론

딥러닝의 성공으로 촉발된 AI 기술은 혁신을 가속화하여 다양한 분야에서 활용 범위를 넓혀 가고 있다. 전통적인 딥러닝 모델은 각 분야의 특정 과제를 목표로 독자적으로 연구, 설계, 개발 및 평가되어 왔으며, 이러한 결과로 개발된 특화 인공지능 모델들은 해당 분야의 성능을 크게 끌어올리고, 학문적 산업적 토대를 마련하는 데 이바지했다. 그런데 이렇게 분야별로 독자적으로 연구하던 흐름이 최근에 바뀌고 있다. 이러한 중심에는 파운데이션 모델(Foundation Model)이 있다.

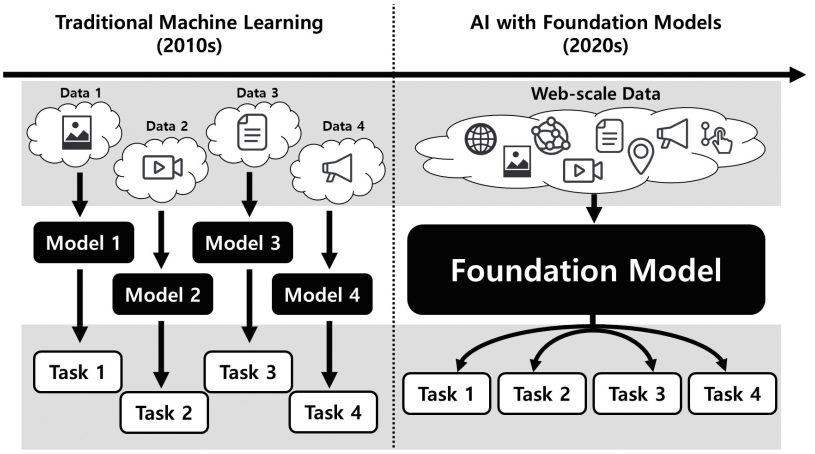
* DOI: https://doi.org/10.22648/ETRI.2025.J.400601
* This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [No.RS-2022-II220124, Development of Artificial Intelligence Technology for Self-Improving Competency-Aware Learning Capabilities].

OPEN 공공누리 출처표시 상업용금지 변경금지 공공저작물 자유이용허락
본 저작물은 공공누리 제4유형
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.
©2025 한국전자통신연구원

파운데이션 모델[1]이란 스탠퍼드대학교 인간 중심 인공지능(HAI: Human-centered Artificial Intelligence) 연구소에서 광범위한 데이터로 훈련된 거대한 인공지능 모델의 등장이 인공지능 연구와 생태계 전반의 패러다임을 바꿀 것이라는 비전을 담아 제시한 개념이다. 이는 두 가지 기술적 사실에 기초하는데, 첫 번째로 거대 언어 모델(LLM: Large Language Model)의 구조적 기반(Architecture, 이하 아키텍쳐)에 해당하는 트랜스포머(Transformer)[2]는 모델의 파라미터와 데이터를 늘릴수록 성능이 향상되고 포화되지 않는 모습이 관찰되었기 때문이다[3]. 두 번째로, 모델의 파라미터와 데이터의 규모가 늘어났을 때, 인공지능 모델이 활용할 수 있는 학습된 사전 학습된 지식이 늘어날 뿐만 아니라, 학습시키지 않은 태스크도 인공지능 모델이 수행할 수 있는 창발적 능력(Emergent Capability)이 발현된다는 사실이 관찰되었기 때문이다[4]. 이를 종합하면 트랜스포머의 모델 규모와 학습 데이터 규모를 늘리면 이전에 비해 비약적으로 범용적인, 다분야에 활용이 가능한 인공지능 모델을 만들 수 있다는 전망을 할 수 있다.

따라서 인공지능 연구의 축이 분야별로 독자적인 연구를 하던 파운데이션 모델의 등장 이전까지의 패러다임에서, 하나의 공통된 파운데이션 모델에서 범용적, 창발적 능력을 최대한 끌어낸 후 개별 분야에 최적화를 추가로 수행하는 연구 패러다임으로 이동하리라는 것이 스탠퍼드 HAI에서 주장한 바라고 생각할 수 있다(그림 1 참고).

이러한 전환은 연구자들이 활동하는 방식과 연구가 조직되는 양상에 깊은 영향을 미쳤다. 파운데이션 모델 학습에는 막대한 비용이 소요되어 연구 주도권이 빅테크에 집중되었으며, 소규모 연구 그룹은 과거처럼 단순히 선도적 성과에 영감을 얻는 수준을 넘어, 그 모델에 직접적으로 종속되는 양상을 보였다. 또한, 다양한 분야의 연구자들이 공통된 기반에서 연구함에 따라 문제의식과 방법론을 공유하게 되는 경향성도 강화되었다. 많은 연구적 주제가 파운데이션 모델을 중심으로 재개편되기도 했다. 연구자들은 파운데이션 모델의 한계를 발견하고 이의 개선에 집중한 연구를 하기도 하였고, 모델의 강력한 범용적, 창발적 성능을 활용한 연구에 집중하기도 하였다. 또한, 일반인들의 AI 활용이 늘어나면서 좀 더 실용적이고 사용자 중심적 관점의 연구가



아이콘 출처 게티이미지뱅크. 무단 전재 및 재배포 금지.

그림 1 AI 연구 패러다임 변화: 태스크별 모델 개발에서 파운데이션 모델 중심의 개발로의 전환

늘어났다.

본고에서는 이러한 연구 동향을 소개하는데, 특히 강조하고 싶은 것은 연구들의 기술적 맥락이다. 본고는 파운데이션 모델의 등장이 변화시킨 환경에서 이뤄진 연구 성과들을 조망하되, 그러한 연구들이 진행된 맥락에 집중하여 개연성을 설명하는 데 노력한다. 그럼으로써 해당 연구들의 중요성에 대한 이해를 돕고, 추후 연구 방향 설정과 동향 예측에 도움을 주고자 한다.

II. 파운데이션 모델 활용 방식 구분

파운데이션 모델의 능력을 활용하는 방법에 관한 초기 연구들은 크게 두 갈래로 구분할 수 있다. 하나는 학습 시점에서 개입하는 방식(Training Time Intervention, 이하 학습 개입 방식)이고, 다른 하나는 추론 시점에서 개입하는 방식(Test Time Intervention, 이하 추론 개입 방식)이다.

학습 개입 방식은 기존 딥러닝 연구의 미세 조정 방식(Fine Tuning)의 연속선상에 있는 접근으로, 사전에 학습된 모델을 재학습하여 목표 태스크의 성능을 높이는 방법을 뜻한다. 파운데이션 모델 연구에서 미세 조정 방식은 곧 지시문 튜닝(Instruction Tuning)[5]으로 확장되었는데, 이는 파운데이션 모델의 학습된 지식을 단순히 방대한 표현학습처럼 사용하는 기존의 관점을 넘어 다수의 태스크 지시문으로 구성된 데이터셋을 통해 모델이 더욱 일반적인 지시 수행 능력을 갖추도록 유도한다는 점에서 혁신적이었다. 기존 미세 조정 방식이 모델의 태스크 맞춤화를 추구했다면, 지시문 튜닝은 다양한 지시문을 활용하여 모델의 범용적 활용성을 높이고 창발적 능력을 더 안정적으로 끌어낼 수 있을 뿐만 아니라 학습 데이터를 목표 태스크에 대한 지시문 수행을 포함하도록 구성하여 목표 태스크의 수행 능력도 향상시킬 수 있다.

추론 개입 방식은 별도의 학습 과정을 거치지 않고, 이미 학습된 파운데이션 모델의 능력을 질의 과정에서 직접 끌어내려는 접근이다. 대표적인 방법은 프롬프팅(Prompting)이며, 특히 프롬프트 내에 목표 태스크 수행의 예시를 통해 모델이 맥락을 파악하도록 유도하는 맥락 내 학습(ICL: In-Context Learning)[4]이 주요한 사례로 제시된다. 이렇게 입력 프롬프트의 설계만으로 LLM이 학습하지 않은 태스크를 수행할 수 있다는 발견은 파운데이션 모델의 창발적 능력에 대해 연구자들이 본격적으로 주목하게 된 중요한 계기가 되었다.

학습 개입 방식과 추론 개입 방식은 서로에 대하여 상대적인 장단점을 가진다. 추론 개입 방식은 별도의 학습을 요구하지 않는다는 점에서 가장 큰 장점이 있다. 이는 현재 파운데이션 모델의 학습 자체가 빅테크 중심으로만 가능하고, 소규모 연구 그룹은 접근하기 어렵다는 현실에서 특히 중요한 의미를 지닌다. 그러나 이러한 접근은 추론 시 연산 자원을 더 소모한다는 단점도 있다. 특히 트랜스포머의 어텐션 모듈이 토큰 수에 따라 제곱적으로 연산량이 증가한다는 점을 고려하면, 프롬프트를 길게 추가하는 방식은 상당한 부담으로 작용할 수 있다. 반대로 학습 개입 방식은 이러한 제약을 피할 수 있으나, 별도의 학습이 필요하다는 점에서 접근성은 떨어진다. 한편, 많은 연구에서 학습 개입 방식이 상대적으로 안정된 성능을 보였다는 결과가 보고되고 있다.

이처럼 파운데이션 모델을 활용하기 위한 초기의 연구는 크게 학습 개입 방식과 추론 개입 방식이라는 두 가지 틀로 이해할 수 있다. 이후 연구들은 이러한 두 범주를 절충하거나 새로운 변형을 제시하면서 발전해 왔다. 따라서 이러한 이분법적 구분은 파운데이션 모델 활용 연구의 기초적 흐름을 이해하고 장단점을 파악하는 데 유용하지만, 오늘날

의 기법들은 이 구분을 넘어서는 다양한 확장과 복합적인 응용으로 이어지고 있다는 사실을 유념해야 한다.

III. 파운데이션 모델의 구조적 한계와 이를 보완하기 위한 기술적 접근

파운데이션 모델의 범용적, 창발적 능력을 활용하기 위한 연구가 다양한 방식으로 진행되면서 파운데이션 모델이 갖는 구조적 제약 또한 점차 뚜렷하게 드러나게 되었다. 여기서 말하는 구조적 제약이란 모델의 학습 방식과 추론 원리에 뿌리를 둔 한계들을 가리킨다. 이는 단순한 구현상의 문제나 자원 제약이 아니라, 모델이 대규모 데이터를 기반으로 학습하고 언어적 생성 과정을 통해 추론을 전개하는 메커니즘에서 본질적으로 파생되는 문제이다. 본 장에서는 이러한 제약 사항 중 일부를 조명하고 이를 해결하기 위한 기술적 접근 방식들을 소개한다.

1. 파운데이션 모델 조정의 어려움

파운데이션 모델의 범용성과 창발성은 거대한 연산 자원과 방대한 학습 데이터를 기반으로 형성된다. 이러한 스케일링 기반 접근법은 필연적으로 막대한 비용을 수반하기 때문에 모델 전체를 필요에 따라 조정하여 사용하는 것은 다소 비현실적인 상태가 되었다.

그러나 필요에 따라 모델이 판단의 근거인 지식을 조정하는 것은 필수적이다. 첫 번째 이유는 빠르게 변화하는 최신 정보를 반영하기 위해서이다. 예를 들어, 'AI 연구자가 노벨상을 받은 적 있느냐'란 질문의 정답은 2024년 이전에는 거짓이었지만 2024년 이후에는 참이 될 것이다. 하지만 모델의 학습이 신속하게 이루어지지 않는다면 모델은 사전 학습된 지식을 바탕으로 거짓이라고 출력할 것이다. 이는 학습데이터의 규모와는 관계가 없고 지식의 반영되는 속도와 관련된 지식 정체 현상(Knowledge Staleness)에 해당한다.

두 번째 이유는 맥락 기반 연결 능력(Contextual Grounding)의 부족 때문이다. 비록 파운데이션 모델은 이미 대규모의 데이터를 통해 학습된 방대한 지식을 바탕으로 다양한 벤치마크에서 뛰어난 성능을 보여주지만, 각 응용 분야에 더 최적화시킬 수 있다. 이는 이른바 '공짜 점심은 없다 이론(No Free Lunch Theorem)'에 기반하는데, 어떤 단일한 모델, 혹은 알고리즘이 모든 문제 영역에서 동시에 최적일 수 없다는 점을 시사한다. 다시 말해, 각 모델은 고유한 귀납적 편향을 내포하고 있으며, 이 편향은 특정 문제에서는 유리하게 작용하지만 다른 문제에서는 제약으로 나타난다. 따라서 파운데이션 모델 역시 전 영역에서 일관된 최적성을 보장할 수 없고, 응용 맥락에 맞춘 추가적인 조정이 불가피하다.

이러한 한계를 보완하기 위한 대표적인 접근 중 하나가 검색 보강 생성 기법(RAG: Retrieval-Augmented Generation)[6]이다. RAG는 모델이 응답을 생성할 때, 모델 파라미터에 내재화된 지식만을 사용하는 것이 아니라 외부 지식 저장소로부터 관련 정보를 검색하여 이를 추론 과정에서 조건부 입력으로 결합한다. 이를 통해 파운데이션 모델은 학습 시점 이후에 등장한 새로운 사실이나 특정 맥락에 특화된 정보를 반영할 수 있게 된다.

RAG가 효과적으로 동작하기 위해서는 LLM이 언제 검색할지를 판단하여야 하고, 목적에 맞는 정보를 찾을 수 있도록 쿼리로 만들 수 있어야 하고, 검색 결과가 목적에 맞는지 판단하고 편집할 수 있어야 한다. 이를 위하여 다양한 RAG 기술 연구가 활발히 진행되고 있다[7,8].

추론 개입 방식에 해당하는 RAG와 달리, 학습 개

입 방식으로 이 문제를 해결하려는 방법도 다수 제시되었다. 대표적으로는 딥러닝 모델의 크기가 증가함에 따라 모든 파라미터를 미세 조정하는 대신 일부 파라미터를 효율적으로 미세 조정하여 성능을 향상하는 파라미터 효율 튜닝(PEFT: Parameter Efficient Fine Tuning)[9]이 있다. 이 계열의 연구 성과 중 최근 가장 널리 쓰이는 방식은 저랭크 적응 기법(LoRA: Low Rank Adaptation)이다. LoRA는 기존의 거대한 가중치 행렬을 그대로 두고, 저차원 투사와 복원을 수행하는 모듈만 추가하여 목표 태스크에 모델을 적응시킨다. 이를 통해 전체 모델을 다시 학습하지 않고도 필요한 보정을 수행하여, 학습 비용과 메모리 사용을 크게 줄일 수 있다.

최근에는 이러한 LoRA를 전문가 혼합 방식(MoE: Mixture of Experts)[10,11] 구조와 결합하려는 시도도 활발하다[12]. MoE는 여러 전문가 모듈을 병렬로 두고, 라우터를 통해 입력 토큰마다 일부 모듈만 선택적으로 활성화하여 추론을 수행하는 방식이다. MoE는 추론 단계에서 각 토큰이 모델 전체를 사용하지 않아 연산 효율을 크게 높일 수 있는데, 이는 FLOPs 관점에서의 효율화로 국한될 뿐, 모델 전체를 GPU 메모리에 대기시켜야 한다는 점에서 메모리 점유율의 관점에서는 효율화시키지 못하는 한계 역시 존재한다. LoRA는 파라미터를 효율화하여 메모리 점유를 줄이기 때문에 이러한 한계를 보완할 수 있다.

2. 긴 추론 경로 유지 능력 부족

LLM의 기반이 되는 트랜스포머는 자기 회귀 생성(Autoregressive Generation) 방식으로 동작한다. 즉, 이미 출력한 토큰들을 조건으로 삼아 이후 어떤 토큰이 가장 적합한지를 판단해 순차적으로 생성하는 구조다. 이러한 방식은 문제 해결을 위한 전체적 계획(Global Plan)을 수립하고 체계적으로 추론하기보다는 국소적 일관성(Local Plausibility)에 따라 다음 출력을 고르는 경향을 강화한다. 이 때문에 모델의 사고 과정은 일관된 체계를 갖추지 못하고, 그 결과물은 단편적이고 즉흥적인 추론의 연쇄로 이어지는 경향성을 보인다. 응답이 길어질수록 이러한 한계는 더욱 두드러져, 오류가 누적되거나 초기 논리의 일관성이 약화되면서 추론이 발산하는 문제가 발생한다.

이러한 긴 추론 경로 유지능력의 부족은 파운데이션 모델로 하여금 논리적 추론(Reasoning)과 같이 체계적이고 일관적인 논리 전개가 요구되는 과제에서 취약점을 드러내게 한다. 따라서 논리적 추론은 현재 LLM의 성능을 평가하고 개선하는 중요한 기준으로 자리 잡고 있다.

이러한 한계를 보완하기 위한 대표적인 접근법이 생각 사슬(CoT: Chain of Thought)이다[13]. CoT는 복잡한 문제를 단순화된 하위 문제 집합으로 분해하고, 모델이 각 단계를 순차적으로 전개하도록 유도함으로써 긴 추론 경로의 사고를 가능하게 한다. 이러한 체계화는 모델이 스스로 하도록 유도할 수도 있고[14], 목표 과제에 맞춰 인간이 설계할 수도 있다[15].

초기 연구에서는 주로 CoT를 프롬프팅을 통하여 모델이 추론 과정을 단계적으로 조정하는 추론 개입 방식으로 사용하였다[13,14]. 그러나 이후에는 지시문 튜닝을 포함한 학습 개입 방식을 포함하여 논리적 추론 문제를 해결하기 위하여 LLM 출력의 체계를 확보하는 전반적인 방법으로 의미가 확장되고 있다[16-18].

또한 CoT처럼 사고 과정의 체계화를 하되, 자기 회귀 생성 방식의 선형성을 극복하고 추론 경로를 다분기적으로 확장하는 방법에 관한 연구가 진행되었다. 예를 들어 생각 트리(ToT: Tree of Thought)[19] 기법은 추론 과정을 분기로 전개하여 다양한 후보해를 평가하는 과정을 포함해 중간 단계 오류에 유

연하게 대응한다. 이는 더 일반화되어 생각 그래프(GoT: Graph of Thought)[20] 기법으로 발전되었다.

ToT와 GoT가 CoT의 목표인 사고 체계의 구조적 확장을 통한 발전이라면 생각 프로그램(PoT: Program of Thought)[21,22]은 표현 방식의 변화를 통한 사고 체계의 발전이라고 볼 수 있다. PoT는 자연어로 생성하는 대신, 이를 프로그래밍 언어의 형식으로 생성하여 추론 과정을 외재화하도록 유도한다. 이는 자연어의 모호함을 제거하고, LLM의 생성 방식과 대비되는 규칙 기반의 결정적(Deterministic) 과정을 활용할 수 있어 더 안정적인 논리적 추론을 수행할 수 있게 하는 장점이 있다.

다양한 CoT 연구를 통하여 LLM이 단순히 최종 답변을 산출하는 것보다, 중간 사고 과정을 외재적으로 유도할 때 더욱 안정적으로 작동한다는 사실이 발견되었다. 또한, CoT를 통하여 LLM이 문제 해결을 위한 사고 과정을 자체적으로 계획할 수 있다는 사실 역시 발견되었다. 이러한 발견은 목표 달성을 위해 중간 단계를 능동적으로 설계하고 연속적 추론을 수행하는 에이전트 AI(Agentic AI)의 개념적 특성과 직접적으로 연결된다. CoT 관련 연구는 LLM을 단순한 응답기가 아닌, 스스로 최종 목적 달성을 위한 사고 과정을 체계화하는 행위자로 해석할 수 있도록 연구자들의 관점을 바꾸었고, 이러한 점은 에이전트 AI의 계획(Planning), 의사결정(Decision Making) 과정으로 자연스럽게 연결되었다. 결과적으로 파운데이션 모델을 토대로 한 에이전트 AI 연구가 빠르게 촉진되었다.

3. 입출력 형태 제한

파운데이션 모델의 연구는 LLM에서 출발했으며, 여전히 많은 파운데이션 모델 개발이 LLM을 중심으로 이루어지고 있다. 가장 큰 이유는 언어 데이터가 인터넷과 문헌을 통해 다른 모달리티보다 압도적으로 대규모로 수집 가능하고, 다른 모달리티에 비해 정보가 압축적이고 체계적으로 담겨 있기 때문이다. 또한, 텍스트는 정규화된 토큰 시퀀스로 변환이 가능하여 대규모 확률 모델링에 적합하고, 동일한 언어 모델 구조가 번역, 요약, 질의응답 등 다양한 다운스트림 태스크를 포괄할 수 있기 때문이다.

그러나 언어 모달리티 역시 여러 가지 고유한 제약을 지닌다. 언어는 기호적·선형적 표현 수단이므로 물리적·다차원적 맥락을 직접적으로 반영하지 못하며, 복잡한 지각 경험을 제한된 어휘와 구문으로 압축하는 과정에서 세부적 정보가 손실된다. 또한, 토큰 단위의 이산적 구조는 연속적인 신호를 표현하기에 부적합하며, 출력 역시 텍스트에 국한되어 실제 행위나 환경과의 직접적 상호작용을 구현하기 어렵다. 이러한 제약은 언어 기반 모델이 갖는 본질적인 한계로, 언어만으로 세계를 충분히 기술하거나 작동하기에는 불완전하다는 점을 암시한다.

따라서 LLM의 입출력에 언어 이외의 모달리티를 연결하는 연구가 활발하게 진행되게 되었다. 멀티모달 거대 언어 모델(MLLM: Multi-modal Large Language Model)은 LLM의 입출력 모달리티를 모두 확장하는 개념을 포괄하지만, 실제 연구와 논의에서는 주로 입력 모달리티의 확장에 초점을 두어 서술되는 경우가 많다. 즉, LLM의 MLLM으로의 확장은 시각, 음성, 비디오와 같은 다양한 입력 신호를 언어와 정렬(Alignment)하여 하나의 모델이 다중 모달리티 정보를 해석할 수 있는 능력을 키우는 연구로 통용될 때가 많다.

언어 이외의 데이터를 LLM에 입력하기 위해서는 두 단계의 연결고리가 필요하다. 첫 번째는 인코더(Encoder) 단계로 데이터로부터 특징 표현을 추출하는 단계이다. 시각 모달리티 연결을 위해서 일반

적으로 비전 트랜스포머(ViT: Vision Transformer) 계열의 인코더를 사용하는데, 대표적으로 CLIP[23], DINOv2[24], EVA-CLIP[25] 등이 여기에 속한다. 비단 시각 모달리티뿐만 아니라 음성 모달리티를 LLM에 연결하는 경우에도 트랜스포머가 일반적으로 사용되고 있다[26,27].

이렇게 인코더를 통과해 특징 표현 벡터로 변환된 언어 외 데이터 정보는 커넥터(Connector)를 통하여 LLM에 연결된다. 이러한 이중 구조를 사용하는 이유는 LLM과 고성능 인코더의 학습이 비용이 많이 들기 때문이다. LLM은 물론이고 위에서 언급한 CLIP, DINOv2, CLIP-EVA 인코더 모두 시각 파운데이션 모델(VFM: Vision Foundation Model)에 해당하여 대규모 데이터로 학습이 이루어진다. 따라서 LLM이나 인코더의 구성요소가 바뀔 때마다 고비용의 재학습을 반복하는 일을 막기 위해서 상대적으로 가벼운 커넥터를 연결하여 호환성을 확보하는 방식을 사용하는 것이다. 현재 다층 퍼셉트론(MLP: Multi Layer Perceptron) 등을 통해 인코더 출력을 투사(Projection)하는 방식[28], 트랜스포머와 학습 가능한 쿼리 토큰을 사용하는 Q-Former 방식[29], 그리고 교차 어텐션(Cross-Attention)층에 별도의 경로를 확보하고 어댑터를 사용하는 방식[30] 등이 널리 쓰인다. 언어 외 데이터 정보는 MLP와 Q-Former를 사용하는 방식의 경우에는 LLM에 입력 토큰 형태로 전달되고, 교차 어텐션 방식은 특징 단계의 융합을 통해 전달한다.

커넥터의 중요성은 연결되는 모달리티와 언어의 표현 간극(Modality Gap)과 상관관계가 높다. 언어와 표현 간극이 큰 시각 모달리티의 경우, 음성과 같이 언어적 정보가 본질적으로 내포된 모달리티보다 커넥터의 고도화가 중요하다.

LLM에서 MLLM으로의 확장은 세계에 존재하는 다양한 형태의 신호를 언어라는 공통 표현 체계 속으로 통합하려는 시도로 이해할 수 있다. 중요한 점은 이러한 통합이 각 모달리티의 고유 특성보다는 LLM의 구조적 제약과 표현 방식에 맞추어 설계된다는 것이다. 예를 들면, 정보가 희소(Sparse)하게 담겨 있는 언어 외 모달리티 데이터를 LLM에 효율적으로 입력하기 위한 연구가 있다. 트랜스포머의 연산량은 토큰 수에 민감하게 반응하고, 따라서 비디오와 같이 정보 대비 고용량의 데이터를 효율적으로 필터링하는 방법에 관한 연구가 활발히 이루어지고 있다[31,32]. 반면, 비디오 행동 이해와 같은 모달리티 고유의 태스크에 관한 부분은 LLM 자체의 창발적 능력을 높이는 CoT 등의 연구의 연장으로 통합되어 가는 경향성이 관찰된다.

LLM의 출력을 비언어 모달리티로 연결하는 예시는 텍스트 기반 영상 생성(T2I: Text to Image) 모델과 시각 언어 액션 모델(VLA model: Vision Language Action model)이 있다. T2I는 입력된 텍스트 지시문을 바탕으로 영상을 생성하는 태스크를 뜻하는데, LLM의 풍부한 언어 이해와 추론 능력을 기반으로 기존 T2I가 처리하지 못했던 복잡한 문맥과 다층적 조건을 반영할 수 있게 되었다. LLM 기반 T2I는 언어 모델에 생성기(Generator)를 디코더로 연결하여 LLM의 출력을 영상의 형태로 바꾸는데, 현재 디퓨전(Diffusion)[33] 방식의 디코더가 가장 일반적으로 사용된다.

T2I 모델이 LLM이 영상을 출력할 수 있게 확장한 모델이라고 한다면 VLA는 로봇의 행동을 제어할 수 있게 하는 모델이다[34]. T2I를 위해 LLM의 출력에 디코더를 연결한 반면, VLA는 로봇의 행동 파라미터를 출력하는 정책 네트워크(Policy Network)를 연결하여 VLM의 출력을 로봇의 행동으로 전환한다. 정책 네트워크의 입력은 언어적 지시문, 시각 정보, 로봇의 상태 정보가 된다. 정책 네트워크는 보통 별도의 트랜스포머로 구현되며 최근에 디퓨전 모

델을 활용하려는 시도도 활발하게 이루어지고 있다.

T2I와 VLA에서 주목할 점 중 하나는 LLM, VLM에서 별도의 디코더를 학습한다는 것이다. VLA를 예시로 들면 로봇의 상태 정보를 토큰화하여 언어적 지시문, 시각 정보와 같은 VLM에 넣어서 액션 파라미터를 하나의 트랜스포머로 뽑아내는 것도 이론상 가능하며, 로봇 제어의 창발적 능력을 발현시킬 가능성도 크다. 하지만 많은 경우 파운데이션 모델을 언어적 지시문과 시각 정보를 임베딩하기 위한 인코더로 사용하고 별도의 정책 네트워크를 따로 구성한다. 이는 파운데이션 모델의 학습 난이도가 높고, 접근권 문제에서 모듈화가 중요한 이슈가 되기 때문이다.

마지막으로, LLM이 현실에 작용할 수 있게 하는 방법에 대한 또 다른 방식으로 ReAct[35]를 언급하고자 한다. ReAct는 행동을 언어로 서술하여 다단계 문제를 해결하는 CoT를 제안하였다. 이는 행동을 언어적 공간에서 기호화함으로써 LLM이 외부 환경과 직접 상호작용할 수 있으며 논리적 추론의 일부로 편입시킬 수 있다는 통찰을 주었다. 이는 곧 언어 지시문을 통하여 검색, 계산, 코드 실행과 같은 외부 모듈을 능동적으로 활용하는 도구 사용(Tool Use) 에이전트 AI에 대한 연구로 일반화되었다. 멀티모달 입력 확장이나 디코더 연결 방식과 달리, LLM의 언어적 표현 자체를 인터페이스로 삼아 세계와 연결하는 새로운 방법론으로 볼 수 있다.

IV. 파운데이션 모델 연구의 발전에 따른 학습, 데이터, 평가 패러다임 변화

Ⅲ장에서 파운데이션 모델의 구조적 한계와 관련된 연구를 소개했다면, 본 장에서는 파운데이션 모델을 통한 발전이 가져온 연구 흐름의 변화를 소개하고자 한다. 파운데이션 모델의 창발성과 범용성은 AI의 응용에서 요구되는 기대 수준을 한 단계 높였으며, 이전에 불가능하던 연구적 시도를 가능하게 하는 토대를 제공하였다. 동시에 일반 사용자층에서도 AI 활용이 급속히 확산되면서, 실제 사용 맥락에서의 요구사항이 연구와 개발의 방향성을 직접적으로 규정하기 시작했다.

이러한 복합적 발전은 AI 생태계 전체를 재편하고 있다. 학습 방식, 데이터 설계, 평가 체계, 그리고 인간과의 상호작용에 이르기까지 기존의 패러다임은 근본적인 변화를 겪고 있으며, 이는 단순한 모델의 성능 향상을 넘어선 새로운 연구 의제들을 촉발하고 있다. 본 장에서는 그러한 연구들을 짚어보고자 한다.

1. 강화학습과 쌍 단위 선호도 비교

대부분의 파운데이션 모델의 학습은 대규모의 데이터를 학습하기 위해 자기 지도 형태로 이루어진다. 그런데 파운데이션 모델을 고도화하는 과정에서 점점 지도 학습이 활용되는 빈도가 높아졌다. 지시문 튜닝을 통해서 지시문 수행 능력을 높이거나 모달리티 간의 데이터쌍을 이용하여 멀티모달 능력을 증진시키는 등 파운데이션 모델의 능력을 확장하고 다운스트림에 적응시키는 학습들에서 추가적인 감독 신호를 이용하는 경우가 점점 많아졌기 때문이다.

그러나 지도 학습은 한계를 빠르게 노출했는데, 첫 번째 이유는 LLM은 태생적으로 생성형 모델이라 정답이 여러 가지 있을 수 있기에 고정된 감독 신호로 감독하는 것이 어렵기 때문이었다. 두 번째 이유는 파운데이션 모델의 성능이 빠르게 상승함에 따라 모델이 의미 있는 학습을 할 수 있는 정적인 지도 학습 데이터의 확보가 점점 어려워졌기 때문이다.

이러한 문제의식은 파운데이션 모델의 고도화를

지도학습 대신 강화학습으로 하려는 시도로 이어졌다. 그 첫 단계라고 할 수 있는 인간 피드백에 의한 강화학습(RLHF: Reinforcement Learning from Human Feedback)[36]에는 두 가지 핵심 디자인이 있는데, 첫 번째는 보상 모델(Reward Model)을 지도 학습으로 학습하고 이 보상 모델을 통하여 LLM을 강화학습 한다는 점이다. 즉 LLM은 인간을 모사한 또 다른 모델인 보상 모델의 학습 신호를 통해 학습이 이루어진다. 두 번째는 보상 모델이 LLM이 생성한 데이터 한 쌍에서 인간의 상대적 선호도를 바탕으로 학습된다는 것이다. 이를 쌍 단위 선호도(Pair-Wise Preference) 비교라고 하는데, 이 방법은 데이터를 LLM이 만든다는 점과 인간의 레이블을 비정밀(Coarse)하게 요구한다는 점에서 학습데이터의 구성 난이도를 혁신적으로 낮출 수 있었다.

RLHF 보상 모델의 학습은 곧 인공지능 피드백에 의한 강화학습(RLAIF: Reinforcement Learning from Artificial Intelligence Feedback)[37]로 빠르게 발전하였다. 즉 학습의 대상이 되는 LLM이 출력한 한 쌍의 답변 간의 선호도를 윤리적 기준을 포함한 다양한 종류의 판단 기준들이 프롬프팅 된 LLM이 판단하여 이를 바탕으로 보상 모델을 학습시키는 방법이다. 이때 답변의 선호도를 판단하는 LLM을 심판(Judge) 모델이라고 한다.

RLAIF에서는 심판 모델을 사용한 학습을 두 가지 트랙으로 사용하였는데, 첫 번째는 심판 모델의 피드백을 받아 곧바로 답변을 재생성하고 이 답변을 교사 학습하는 방법이다. 이때 심판 모델이 언어 형태의 피드백을 줄 수 있기 때문에 LLM은 그 피드백을 받아 곧바로 더 나은 교사 학습 데이터를 생성할 수 있다는 점이 중요하다. 두 번째는 RLHF와 같이 심판 모델을 사용하여 보상 모델을 학습한 후 목표 LLM을 강화학습하는 것이다. 이러한 두 가지 학습 방법을 병용하는 이유는 LLM의 학습이 충분히 진행되지 않은 상태에서 강화학습은 불안정하므로 교사 학습을 통하여 빠르게 학습을 안정화시키기 위해서다.

한편 RLHF, RLAIF에서 강화학습의 불안정성을 해결하고자 보상 모델을 없애고 선호도 비교 요소만 남겨 쌍 단위 교사 학습 형태로 학습 방식을 바꾼 직접 선호 최적화(DPO: Direct Preference Optimization)[38] 방식이 제안되었다. DPO는 손실 함수의 디자인을 통하여 학습이 정책 그래디언트 업데이트를 근사할 수 있도록 하여 보다 안정적인 학습이 가능하게 하였다.

앞서 살펴본 RLHF에서 RLAIF를 걸쳐 DPO로 이어지는 발전 과정은 RLHF의 외부 학습 신호와 쌍 단위 선호도라는 특성을 유지한 채, 보상 모델의 구현을 좀 더 효율화하려는 방향성의 노력이다. 이러한 방향성과는 다르게 최근 세계 모델(World Model) 철학을 파운데이션 모델과 결합하려는 노력이 점점 늘어나고 있다[39,40].

세계 모델은 강화학습에선 오랫동안 사용하던 개념인데, 환경을 시뮬레이션하여 에이전트의 행동 결과를 예측할 수 있도록 하는 모델을 뜻한다. 이러한 연구가 점점 확산되는 이유는 크게 두 가지로 볼 수 있는데, 첫 번째 이유는 어떠한 태스크들은 쌍 비교나 언어 기반의 심판 모델이 태생적으로 어울리지 않을 수 있기 때문이다. 예컨대 로봇 제어와 같이 물리적 연속 공간에서의 행동 결정을 요구하는 태스크는 쌍 단위 비교나 언어논리로 감독하고 피드백하기 어렵다. 이런 경우에는 RLHF 계열의 학습보다는 프로그램이나 모델에 의한 시뮬레이션이 개입하는 학습이 더 적합하다고 쉽게 예측할 수 있다.

두 번째는 강화학습의 보상 모델을 내재화시켰을 때 자가 개선(Self-Improving)이 가능할 것이라는 파운데이션 모델에 대한 기대감 때문이다. 즉 어떠한 연구는 학습하고자 하는 LLM과 세계 모델을 같은

모델로 사용하는데, 이러한 경우 학습하고자 하는 LLM이 시뮬레이션 능력을 얻어 행동 결과를 예측하고 평가할 수 있는 능력을 같이 갖춰 메타인지적 능력을 키운 AI가 될 수 있을 것이라는 기대와 함께 다양한 연구가 시도되고 있다.

2. 데이터 중심 인공지능

딥러닝 연구는 전통적으로 모델 중심(Model-Centric) 접근에 치중해 왔다. 컨볼루션 신경망(CNN: Convolution Neural Network) 시절부터 새로운 네트워크 구조, 학습 기법, 정규화 전략 등 모델 설계의 혁신이 성능 향상을 주도했고, ImageNet과 같은 고정된 벤치마크가 사실상 표준으로 자리 잡으면서 데이터 자체를 개선하거나 재구성하는 연구는 상대적으로 소홀히 다뤄졌다. 이러한 흐름은 LLM 시대에도 이어져, 학계는 데이터의 품질보다는 모델이 데이터의 불완전성을 극복하도록 설계하는 데 집중해 왔다.

그러나 지시문 튜닝과 같은 학습 데이터 디자인이나, ICL처럼 추론 단계에서 데이터를 맥락으로 제공하는 방법을 통해 큰 성능 향상이 가능하단 사실을 발견한 연구자들은 데이터 중심(Data-Centric) 연구에 집중하기 시작하였다.

특히 CLIP[23]의 성능을 재현하는 과정에서 연구자들은 CLIP의 일반화 성능이 모델의 구조나 학습 방법에 있는 것이 아닌 데이터에 있다는 사실을 발견하였다[41]. 그러나 빅테크 기업들이 성능의 핵심인 데이터와 그 수집 과정을 불투명하게 다루고, 모델과 학습 방법만을 제한적으로 공개함으로써 재현 가능성과 연구 생태계의 민주성을 약화시킨다는 주장이 제기되었다. 이러한 문제의식에서 나온 대표적 성과가LAION[42]으로 대규모 이미지-텍스트 데이터를 공개해 누구나 활용할 수 있도록 함으로써 연구의 개방성과 민주화를 촉진하였다.

LAION 연구진들은 이러한 철학을 진전시켜 모델이 아닌 데이터를 평가하는 Datacomp[43]를 선보였다. 모델을 평가할 때 데이터를 고정하는 것과 반대로, Datacomp에서는 모델의 아키텍쳐와 학습 방법을 고정한 채, 학습 데이터를 변화시켜서 학습 결과를 비교한다. 이로써 모델의 성능 차이의 독립변인을 데이터로 제한하고 데이터를 어떻게 구성해야 성능과 일반화 능력을 끌어낼 수 있는지를 체계적으로 분석할 수 있는 틀을 제공한다.

데이터 중심 연구의 또 다른 대표적인 성과로는 Molmo&Pixmo[44]를 들 수 있다. 해당 연구는 데이터의 큐레이션을 통하여 많은 상용 파운데이션 모델의 성능을 추월하였다. 저자들은 인터넷에서 크롤링하는 데이터가 우연적일 뿐만 아니라 인간의 캡셔닝 패턴의 편향성을 반영한다고 한다. 가령 사람들이 SNS(Social Network Service)에 사진을 올릴 때 물체 간의 위치 관계 정보를 포함하는 캡셔닝을 할 가능성은 작고, 이를 바탕으로 학습할 경우 모델은 공간적 관계에 관한 서술이 뜻하는 바가 무엇인지 학습하지 못할 가능성이 높다는 것이다. 이러한 문제를 해결하기 위하여 저자들은 사람들이 직접 캡셔닝하게 하되, 타이핑하게 시키는 대신 대신 녹음을 하도록 한 후 TTS(Text To Speech)로 텍스트화하여 캡셔닝의 편의를 높여주거나 구체적인 가이드라인을 주는 등 학습 데이터의 품질을 높이는 여러 방법을 고안하여 시도하였다.

마지막으로, 기존 데이터셋을 리팩토링하는 연구가 급격히 증가하고 있다. 이러한 흐름은 주로 지시문 튜닝과 결합되는데, 지시문 튜닝은 미세 조정을 위한 데이터를 디자인한다는 점에서 본질적으로 데이터 중심 접근이다. 이러한 흐름은 최근 비약적으로 성능이 향상된 상업적 모델을 활용하여 양질의 리팩토링이 가능해졌기 때문에 주목받고 있다.

3. 벤치마킹 변화

파운데이션 모델의 등장에 따라 벤치마크 방식도 재편할 필요성이 생겼다. 기존의 벤치마크는 단일 태스크에 대한 성능 평가를 위해 디자인되었는데 이는 범용성과 창발성에 집중하는 파운데이션 모델의 평가에는 근본적으로 맞지 않았기 때문이다.

파운데이션 모델의 능력을 종합적으로 평가하기 위한 초창기 시도는 여러 태스크를 종합적으로 측정하는 방식에 의존하였다. 이러한 연구의 예시로는 MMLU[45]가 있는데, 인문학, 사회 과학, 자연과학 등 다양한 분야의 질문을 시험하는 형식으로 구성되었다. 이는 파운데이션 모델의 평가가 특정 태스크의 수행 능력 평가에서 다양한 분야의 지식을 시험하는 관점으로 옮겨갔다는 걸 시사한다.

그러나 단순히 지식의 폭을 시험하는 접근은 모델의 실제 활용 가능성을 평가하는 데에는 불충분했다. 파운데이션 모델이 일반인에게까지 널리 쓰이면서 모델의 다면적인 평가 요소가 필요하다는 문제의식이 대두되었다. MMLU 계열의 평가는 다양한 문제를 시험하지만 평가의 요소는 정확도로 제한되었다. 이러한 문제의식 속에서 등장한 HELM[46]은 학문적 분류보다는 기사 요약, 번역, 프로그래밍, 연속된 대화 등 실사용 시나리오를 중심으로 평가를 재편하고, 정확도뿐만 아니라 공정성, 안정성, 효율성 등 다차원 지표를 동시에 제시함으로써 모델의 품질을 총체적으로 파악하려는 시도를 제안하였다.

HELM은 파운데이션 모델을 실제 사용자 중심(Human Centric)으로 평가하기 위한 큰 출발점이었으나, 초창기 연구인 만큼 한계가 있었다. 먼저 공정성, 안정성 등 인간적 가치 평가가 필요한 곳에는 인간의 평가를 통해서 모델을 채점할 수밖에 없었다는 것이다. LLM의 답변이 인간의 기준을 만족시키는지 확인하는 것은 비용적으로 많이 들고 일관성도 부족했다.

이러한 문제를 해결하기 위하여 평가에 필요한 인간의 노력을 줄이는 연구들이 진행되었다. 대표적인 것은 Chatbot Arena[47], LM Arena[48]인데, 이 연구들은 하나의 질의에 대해 LLM에서 나온 두 결과물, 혹은 서로 다른 LLM에서 나온 두 결과물을 사용자에게 제시한 후, 사용자의 피드백을 받아 선호도를 평가할 수 있는 시스템을 만들었다. 이는 앞서 설명한 RLHF나 DPO 등 쌍 단위 선호도 비교 기반 학습 방식과 쉽게 연결되었기 때문에 서로 상승효과를 일으키며 발전할 수 있었다. 또한, 이런 방향의 연구가 발전해서 과연 인간이 높은 선호도를 보이는 답변은 어떠한 요소를 갖는가에 관한 연구로 확장되기도 하였다[49].

또한, HELM의 데이터가 정적이라는 점도 문제가 되었다. 파운데이션 모델은 점점 에이전트 AI적 요소가 강화되고 있는데 HELM의 시나리오와 지표는 고정적이기 때문에 이러한 능력을 평가하기에 적합하지 않았다. 이에 따라 도구를 사용하고, 환경과 상호작용을 하며, 다단계 추론을 거쳐 최종 결과를 산출하는 능력을 시험하는 새로운 벤치마크 GAIA[50]가 등장하여 파운데이션 모델의 에이전트적 성능을 평가할 수 있는 틀을 마련하였다.

마지막으로 새로운 벤치마크가 공개되는 빈도가 폭증하고 있다는 것을 언급하고 싶다. 앞서 설명하였듯이 상용 파운데이션 모델을 활용해서 양질의 데이터셋 리팩토링이 가능해졌고, 이로 인해 자신의 문제의식과 연결되는 벤치마크를 새로 구성하여 공개하는 연구 방식이 이전에 비해 크게 보편화되었다. 따라서, 기존에 공개된 벤치마크에 의존하여 제한되던 문제설정과 연구설계를 연구자 개개인이 좀 더 자유롭게 설정할 수 있게 되었다.

V. 결론

파운데이션 모델은 단일 알고리즘이나 기법을 넘어선 거대한 프로젝트이며, 그 개발은 소수의 빅테크 기업이 주도하고 있다. 실제로 모델을 직접 소유하거나 학습할 수 있는 주체는 극히 제한적이며, 따라서 각 연구 주체의 연구의 폭에 여러 가지 제약이 따른다. 또한, 하나의 연구 성과가 공통 기반인 파운데이션 모델의 성능과 성격을 바꾸어 학계와 산업 전반에 걸쳐 광범위한 영향을 미친다. 이러한 환경에서는 협력과 조율, 포지셔닝이 무엇보다 중요하다.

따라서 이제는 단순히 새로운 기술이 등장했다는 사실을 아는 것만으로는 충분하지 않다. 다른 연구자들이 무엇에 주목하고, 왜 그러한 문제에 집중하는지를 파악하는 것이 점점 필수적으로 요구된다. 본고는 이러한 문제의식을 바탕으로, 기술을 단순히 나열하기보다는 그 배경과 맥락을 드러내고자 하였다. 본고가 파운데이션 모델 생태계를 이해하고 연구와 개발 방향을 모색하는 데 도움을 줄 수 있기를 희망한다.

독자적인 파운데이션 모델 개발은 이런 맥락에서 중요한 의미를 갖는다. 단순히 기술을 확보하는 차원을 넘어서, 국가 차원의 연구 자율성을 확보하고 글로벌 기업들이 주도하는 생태계 안에서도 국내 연구자들이 스스로 연구 방향을 정할 수 있는 토대를 만드는 작업이기 때문이다. 한국전자통신연구원은 NC AI 컨소시엄 참여를 통해 국내 AI 연구 역량을 쌓고, 국제 경쟁 속에서 우리만의 독창적인 성과를 만들어내는 데 기여하고자 한다.

또한, 파운데이션 모델의 뛰어난 범용성과 확장성은 그 강력한 성능을 활용하고 한계를 극복하기 위한 다양한 보강 장치의 개발로 이어졌다. 앞으로도 이러한 연구가 가속화되고 파운데이션 모델은 다양한 보강 장치를 보유하게 될 것이 분명하다. 그러나 이러한 상황에서, 인공지능 모델이 어떤 맥락에서 어떤 방법을 선택하고 결합해야 하는지 판단할 수 있도록 하는 문제는 여전히 어려운 과제로 남아 있다. 이러한 문제를 해결하는 데 필요한 연구가 곧 AI의 메타인지적 능력에 관한 연구이며, 이는 파운데이션 모델이 단순한 성능 향상을 넘어 신뢰 가능한 지능으로 발전하는 데 핵심적 토대가 될 것이다. 한국전자통신연구원 시각지능연구실에서는 이러한 비전을 갖고 '스스로 학습역량을 인지하고 활용하여 적정한 결과를 제공하는 인공지능 기술 개발' 과제를 수행하고 있다.

용어해설

Chain-of-Thought(CoT) 복잡한 문제 해결을 위해 모델이 중간 추론 과정을 단계별로 생성하도록 유도하여 답변의 정확도와 설명 가능성을 높이는 기법

In-Cntext Learning(ICL) 모델이 사전 학습된 파라미터를 수정하지 않고, 입력 프롬프트에 제공된 예시나 지시문을 활용하여 새로운 과제를 수행하는 능력

Low Rank Adaptation(LoRA) 대규모 언어모델의 사전학습 가중치를 고정한 상태에서, 각 선형 변환이나 어텐션 가중치에 저차원 행렬을 추가해 필요한 부분만 학습하는 방식으로, 파라미터 수와 연산량을 크게 줄이면서 효율적인 미세 조정을 가능하게 하는 기법

Mixture of Expert(MOE) 여러 개의 전문가 모듈 중 입력에 따라 일부만 선택적으로 활성화하여 연산을 수행하는 구조로, 모델 용량은 크게 확장하면서도 실제 추론 시 사용되는 연산량을 줄여 효율성을 확보하는 기법

Retrieval-Augmented Generation(RAG) 외부 지식 저장소에서 관련 정보를 검색해 모델 입력에 결합함으로써 모델이 사전 학습에 없는 최신 정보나 구체적 사실을 반영하여 더 정확하고 신뢰성 있는 응답을 생성하도록 하는 기법

강화학습 강화학습 환경과 상호작용을 하는 에이전트가 보상 신호를 최대화하도록 행동을 학습하는 방법으로, 탐험과 활용의 균형을 통해 최적의 정책을 습득하는 기계학습 기법. 주로 명시적 정답이 없는 상황에서 장기적인 성과를 최적화하기 위해 사용되며, 게임 플레이, 로보틱스 제어, 추천 시스템 등에서 효과적

교사 학습 입력과 정답 레이블이 짝지어진 데이터로 모델을 학습시켜, 주어진 예시를 따라 정답을 예측하도록 만드는 지도학습 기법

자기 교사 학습 학습 정답 레이블이 없는 데이터로부터 모델 스스로 학습 과제를 생성하고 이를 통해 표현을 학습하는 기법

자기회귀 생성 이전에 출력으로 생성된 토큰을 조건으로 다음 토큰을 순차적으로 예측하여 시퀀스를 만들어내는 생성 기법

모달리티 인간이나 기계가 정보를 표현·지각·처리하는 감각적 양식으로, 텍스트·음성·영상·센서 데이터와 같이 서로 다른 형태의 입력 또는 출력 채널을 의미함

쌍 대비 선호도 비교 두 개의 출력 결과를 제시하고 그중 어느 쪽이 더 바람직한지 평가자가 선택하도록 하는 방식의 피드백 수집 기법

미세튜닝 사전학습 된 모델을 특정 과제나 도메인에 맞추기 위해 추가 데이터로 파라미터를 조정하는 기법

에이전트 AI 에이전트 AI 사용자의 지시나 목표를 받아들여 환경과 상호작용을 하며 작업을 스스로 계획·수행하는 목표 지향적 인공지능 시스템. 단순 질의응답을 넘어 도구 활용, 의사결정, 멀티스텝 추론 등을 통해 복합적인 과제를 해결하도록 설계됨

지시문 튜닝 지시문 튜닝 모델이 사용자의 명령이나 요청을 더 잘 따르도록, 지시문과 그에 따른 응답 데이터를 활용해 사전학습된 모델을 추가로 학습시키는 기법

파운데이션 모델 대규모 데이터와 연산 자원으로 사전학습되어, 이후 다양한 다운스트림 과제에 맞게 적응할 수 있는 범용 인공지능 모델

프롬프팅 언어모델에 원하는 출력이나 동작을 유도하기 위해 입력 형태나 문구를 설계·제공하는 기법

참고문헌

[1] R. Bommasani et al., "On the opportunities and risks of foundation models," arXiv preprint, 2021. doi: 10.48550/arXiv.2108.07258

[2] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst., (Long Beach, CA, USA), Dec. 2017.

[3] J. Kaplan et al., "Scaling laws for neural language models," arXiv preprint, 2020. doi: 10.48550/arXiv.2001.08361

[4] T.B. Brown et al., "Language models are few-shot learners," in Proc. Adv. Neural Inf. Process. Syst., Dec. 2020, pp. 1877-1901.

[5] J. Wei et al., "Finetuned language models are zero-shot learners," arXiv preprint, 2021. doi: 10.48550/arXiv.2109.01652

[6] G. Izacard and G. Edouard, "Leveraging passage retrieval with generative models for open domain question answering," arXiv preprint, 2020. doi: 10.48550/arXiv.2007.01282

[7] Z. Wang et al., "Learning to filter context for retrieval-augmented generation," arXiv preprint, 2023. doi: 10.48550/arXiv.2311.08377

[8] M. Glass et al., "Re2G: Retrieve, Rerank, Generate," arXiv preprint, 2022. doi: 10.48550/arXiv.2207.06300

[9] N. Houlsby et al., "Parameter-efficient transfer learning for NLP," in Proc. Int. Mach. Learn., (Long Beach, CA, USA), Jun. 2019.

[10] R.A. Jacobs et al., "Adaptive mixtures of local experts," Neural Comput., vol. 3, no. 1, 1991, pp. 79-87.

[11] W. Fedus et al., "Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity," J. Mach. Learn. Res., vol. 23, no. 1, 2022, pp. 1-39.

[12] S. Dou et al., "Loramoe: Revolutionizing mixture of experts for maintaining world knowledge in language model alignment," arXiv preprint, 2023. doi: 10.48550/arXiv.2312.09979

[13] J. Wei et al., "Chain-of-thought prompting elicits reasoning in large language models," in Proc. Adv. Neural Inf. Process. Syst., (New Orleans, LA, USA), Nov. 2022, pp. 24824-24837.

[14] T. Kojima et al., "Large language models are zero-shot reasoners," in Proc. Adv. Neural Inf. Process. Syst., (New Orleans, LA, USA), Nov. 2022, pp. 22199-22213.

[15] H. Shao et al., "Visual cot: Unleashing chain-of-thought reasoning in multi-modal language models," arXiv preprint, 2024. doi: 10.48550/arXiv.2403.16999

[16] H. Liu et al., "Logicot: Logical chain-of-thought instruction-tuning," arXiv preprint, 2023. doi: 10.48550/arXiv.2305.12147

[17] L.H. Li et al., "Symbolic chain-of-thought distillation: Small models can also "think" step-by-step," arXiv preprint, 2023. doi: 10.48550/arXiv.2306.14050

[18] X. Zhang et al., "Chain of preference optimization: Improving chain-of-thought reasoning in LLMs," in Proc. Adv. Neural Inf. Process. Syst., (Vancouver, BC, Canada), Dec. 2024, pp. 333-356.

[19] S. Yao et al., "Tree of thoughts: Deliberate problem solving with large language models," in Proc. Adv. Neural Inf. Process. Syst., (New Orleans, LA, USA), Dec. 2023, pp. 11809-11822.

[20] M. Besta et al., "Graph of thoughts: Solving elaborate problems with large language models," in Proc. AAAI Conf. Artif. Intell., (Vancouver, BC, Canada), Feb. 2024, pp. 17682-17690.

[21] W. Chen et al., "Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks," arXiv preprint, 2022. doi: 10.48550/arXiv.2211.12588

[22] L. Gao et al., "Pal: Program-aided language models," in Proc. Int. Conf. Mach. Learn., (Honolulu, HA, USA), Jul. 2023, pp. 10764-10799.

[23] A. Radford et al., “Learning transferable visual models from natural language supervision,” in Proc. Int. Conf. Mach. Learn., Jul. 2021, pp. 8748-8763.

[24] M. Oquab et al., “Dinov2: Learning robust visual features without supervision,” arXiv preprint, 2023. doi: 10.48550/arXiv.2304.07193

[25] Q. Sun et al., “Eva-clip: Improved training techniques for clip at scale,” arXiv preprint, 2023. doi: 10.48550/arXiv.2303.15389

[26] A. Radford et al., “Robust speech recognition via large-scale weak supervision,” in Proc. Int. Conf. Mach. Learn., (Honolulu, HA, USA), Jul. 2023, pp. 28492-28518.

[27] Y. Gong et al., “Ast: Audio spectrogram transformer,” arXiv preprint, 2021. doi: 10.48550/arXiv.2104.01778

[28] H. Liu et al., “Visual instruction tuning,” in Proc. Adv. Neural Inf. Process. Syst., (New Orleans, LA, USA), Dec. 2023, pp. 34892-34916.

[29] J. Li et al., “Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models,” in Proc. Int. Conf. Mach. Learn., (Honolulu, HA, USA), Jul. 2023, pp. 28492-28518.

[30] J.B. Alayrac et al., “Flamingo: a visual language model for few-shot learning,” in Proc. Adv. Neural Inf. Process. Syst., (New Orleans, LA, USA), Nov. 2022, pp. 23716-23736.

[31] Z. Wang et al., “Videotree: Adaptive tree-based video representation for llm reasoning on long videos,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (Nashville, TN, USA), June. 2025, pp. 3272-3283.

[32] S. Han et al., “Videoespresso: A large-scale chain-of-thought dataset for fine-grained video reasoning via core frame selection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., (Nashville, TN, USA), June. 2025, pp. 26181-26191.

[33] J. Ho et al., “Denoising diffusion probabilistic models,” in Proc. Adv. Neural Inf. Process. Syst., Dec. 2020, pp. 6840-6851.

[34] A. Brohan et al., “Rt-2: Vision-language-action models transfer web knowledge to robotic control,” in Proc. Mach. Learn. Res., (Atlanta, GA, USA), Nov. 2023, pp. 2165-2183.

[35] S. Yao et al., “React: Synergizing reasoning and acting in language models,” in Proc. Int. Conf. Learn. Represent., (Kigali, Rwanda), May. 2023.

[36] D.M. Ziegler et al., “Fine-tuning language models from human preferences,” arXiv preprint, 2019. doi: 10.48550/arXiv.1909.08593

[37] Y. Bai et al., “Constitutional AI: Harmlessness from AI feedback,” arXiv preprint, 2022. doi: 10.48550/arXiv.2212.08073

[38] R. Rafailov et al., “Direct preference optimization: Your language model is secretly a reward model,” in Proc. Adv. Neural Inf. Process. Syst., (New Orleans, LA, USA), Dec. 2023, pp. 53728-53741.

[39] S. Hao et al., “Reasoning with language model is planning with world model,” arXiv preprint, 2023. doi: 10.48550/arXiv.2305.14992

[40] Y. LeCun, “A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27,” Open Review, 2022, pp. 1-62.

[41] H. Xu et al., “Demystifying clip data,” arXiv preprint, 2023. doi: 10.48550/arXiv.2309.16671

[42] C. Schuhmann et al., “Laion-5b: An open large-scale dataset for training next generation image-text models,” in Proc. Adv. Neural Inf. Process. Syst., (New Orleans, LA, USA), Nov. 2022, pp. 25278-25294.

[43] S.Y. Gadre et al., “Datacomp: In search of the next generation of multimodal datasets,” in Proc. Adv. Neural Inf. Process. Syst., (New Orleans, LA, USA), Dec. 2023, pp. 27092-27112.

[44] M. Deitke et al., “Molmo and pixmo: Open weights and open data for state-of-the-art multimodal models,” arXiv preprint, 2024. doi: 10.48550/arXiv.2409.17146

[45] D. Hendrycks et al., “Measuring massive multitask language understanding,” arXiv preprint, 2020. doi: 10.48550/arXiv.2009.03300

[46] P. Liang et al., “Holistic evaluation of language models,” arXiv preprint, 2022. doi: 10.48550/arXiv.2211.09110

[47] W.L. Chiang et al., “Chatbot arena: An open platform for evaluating llms by human preference,” in Proc. Int. Conf. Mach. Learn., (Vienna, Austria), Jul. 2024, pp. 8359-8388.

[48] https://lmarena.ai/

[49] L. Dunlap et al., “Vibecheck: Discover and quantify qualitative differences in large language models,” arXiv preprint, 2024. doi: 10.48550/arXiv.2410.12851

[50] G. Mialon et al., “Gaia: a benchmark for general AI assistants,” in Proc. Int. Conf. Learn. Represent., (Vienna, Austria), May. 2024.