FOCUS 1

신뢰성 있는 사회 인프라의 AX를 위한 뉴로심볼릭 에이전틱 AI 컴퓨팅

• **김태호** | 정보통신기획평가원 SW PM | taehokim@iitp.kr



1. 뉴로심볼릭 AI의 등장과 필요성

AI가 질문에 답하는 수준을 넘어, 목표를 이해하고 계획을 세워 외부 도구로 실제 업무를 수행하는 에이전틱 AI와 로봇, 자동차, 산업설비처럼 물리 환경에서 행동하는 피지컬 AI로 확대되고 있다. 이제 AI에게 요구되는 것은 더 좋은 답변이 아니라 신뢰할 수 있는 실행이다.

그러나, 딥러닝과 파운데이션 모델만으로는 신뢰를 구현하기 어렵다. 딥러닝은 확률적으로 동작하므로 그럴듯하지만 틀린 결과를 낼 수 있고, 반드시 지켜야 하는 제한 조건을 항상 지킨다고 보장하기 어렵다. 챗봇의 오답은 다시 물으면 되지만, 금융 에이전트의 잘못된 송금이나 원자력 발전소의 운영 에이전트가 일으키는 사고는 되돌릴 수 없는 피해가 된다. 사회 인프라에 관련된 분야는 성능 요구사항을 만족하는 것만으로는 부족하며, 실행의 근거와 조건을 확인할 수 있어야 한다.

이러한 한계를 보완하는 접근이 '뉴로심볼릭 AI'이다. 뉴로심볼릭 AI는 데이터에 의한 뉴럴네트워크와 지식을 표현하는 심볼릭 시스템, 두 방식의 AI의 장점을 결합한다. 뉴럴네트워크는 이미지, 언어, 센서 데이터와 같은 다양한 비정형 정보를 인식하고 새로운 후보를 생성한다. 그리고, 심볼릭 시스템은 명시적으로 표현된 지식, 규칙, 논리, 제약조건을 이용하여 추론하고 결과를 검증한다. 요약하면, 뉴럴네트워크는 '가능한 답'을 생성하고, 심볼릭 시스템은 그 답이 '허용 가능한가'를 판단하는 구조로 활용될 수 있다. 이를 에이전트에 적용한 것이 이 글의 '뉴로심볼릭 에이전틱 AI'이며, 뉴럴네트워크와 심볼릭 시스템이 상호보완적으로 동작한다.

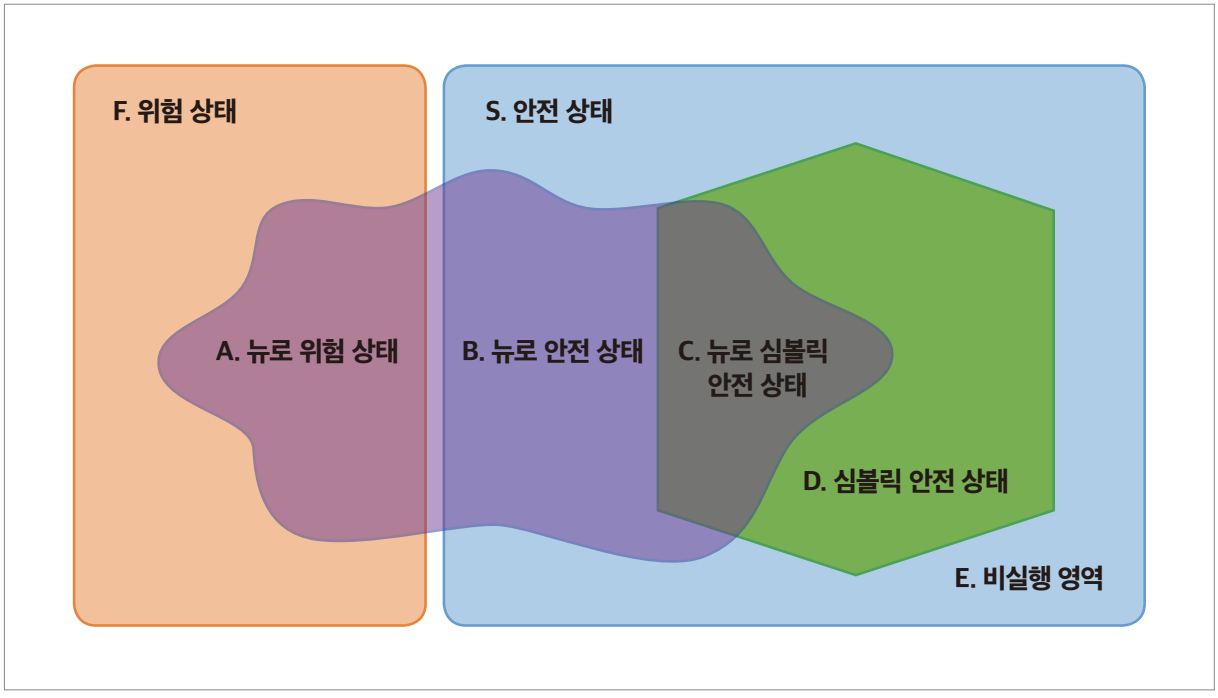
[그림 1] 뉴럴네트워크, 심볼릭 시스템, 뉴로심볼릭 AI의 관계



자율주행 자동차가 공사 구간을 지나고 있는데, 신호등은 녹색이지만 사람은 정지 수신호를 보낸다고 가정하자. 뉴럴네트워크는 신호등, 수신호, 공사 표지를 인식해 서행 통과 또는 정지를 계획한다. 도로교통법은 교통안전시설의 신호와 경찰의 지시가 다르면 경찰을 따르도록 하고 있다. 하지만, 민간 신호수에 대해서는 규정이 모호하기 때문에 상황별 판단이 필요하고, 결정할 수 없는 상황은 판단 불가로 하여 정지하거나 인간의 결정을 기다려야 한다. 이러한 내용은 명확한 규칙으로 정의되지 않으면 자율주행 자동차는 혼란에 빠질 것이다. 또한, 이러한 규정이 개정될 때 시간이 오래걸리거나 결과를 예측할 수 없는 것이 아니라, 빠르고 확실하게 적용되어야 한다.

이 상황은 뉴럴네트워크 기반의 딥러닝만으로는 해결이 어렵다. 첫째, 규정 준수를 보증할 수 없다. 신호와 신호수의 지시가 충돌하는 상황은 드물어 학습 사례가 부족하고, 대부분 정지하더라도 항상 정지한다는 보증이 불가능하다. 둘째, 판단 근거를 확인하고 갱신하기 어렵다. 규칙이 가중치에 암묵적으로 녹아 있으면 정지 이유를 법 조항과 대조할 수 없고, 법이 바뀌면 모델을 다시 학습하고 시험해야 한다. 이는 시간이 소요된다. 셋째, 판단 불가능한 상황을 구분하기 어렵다. 뉴럴네트워크 모델에서 신뢰도(confidence)는 과신되는 경향이 있으며, 엣지 케이스에 대한 처리가 어렵다. 이는 학습을 더 한다고 완전하게 해결되는 문제가 아니라 학습 방식의 구조적 한계이다. 최근 CoT (Chain-of-Thought) 등으로 원인을 설명할 수 있고 정확도가 향상되고 있지만, 보증이 불가능하다는 것은 변하지 않는다.

[그림 2] 딥러닝 실행 영역과 심볼릭 영역



[그림 2]는 시스템의 상태를 크게 안전 영역(S, Safe)과 위험 영역(F, Failure)으로 구분한다. 여기서 안전 영역은 사고가 발생하지 않는 상태를 의미하며, 반드시 유용하거나 정상적인 상태만을 뜻하지는 않는다. 예를 들어 자동차는 정지해 있을 때 가장 안전할 수 있지만, 고장으로 움직이지 못하는 상태 역시 안전 영역에 포함될 수 있다.

보라색의 A, B, C 영역은 뉴럴네트워크가 판단하거나 행동을 생성할 수 있는 상태를 나타낸다. 반면 초록색의 C, D 영역은 규칙과 논리 등을 이용해 심볼릭 시스템이 상태를 표현하고 안전성을 검증할 수 있는 영역을 의미한다.

이 중 A 영역은 뉴럴네트워크가 만들어낼 수 있지만 위험 영역(F)에 속하는 상태이다. 따라서 실제 시스템에서는 이러한 상태로 진입하지 않도록 차단해야 한다.

B 영역은 실제로는 안전 영역(S)에 속하지만, 심볼릭 시스템이 충분히 표현하거나 검증하지 못하는 상태이다. 즉 안전할 가능성은 있지만 이를 명시적으로 입증하기 어렵다. 따라서 원자력, 자동차 등 높은 안전성이 요구되는 시스템에서는 보수적으로 실행을 제한하거나 추가적인 검증을 수행할 필요가 있다.

C 영역은 뉴럴네트워크가 생성할 수 있으면서 동시에 심볼릭 시스템이 안전성을 검증할 수 있는 영역이다. 따라서 신뢰성 있는 뉴로심볼릭 AI가 우선적으로 동작해야 하는 영역이라고 볼 수 있다.

D 영역은 심볼릭 시스템이 안전하다고 표현하거나 검증할 수 있지만, 현재 뉴럴네트워크가 실제 행동으로 생성하지 않는 상태를 의미한다. 이러한 영역은 향후 뉴럴네트워크의 학습과 탐색 범위를 확대하는 데 활용될 수 있다.

이상적인 뉴로심볼릭 AI 시스템에서는 뉴럴네트워크가 생성하는 상태가 안전 영역(S)을 벗어나지 않아야 하며, 가능한 한 심볼릭 시스템이 검증할 수 있는 영역 안에서 동작해야 한다. 이를 위해 뉴럴네트워크는 안전하면서도 유용한 상태의 범위를 넓히고, 심볼릭 시스템은 실제 안전 영역을 최대한 폭넓게 표현하고 검증할 수 있어야 한다.

뉴로심볼릭 AI는 딥러닝을 대체하는 기술이 아니라, 딥러닝의 학습, 추론, 생성 능력에 지식, 논리, 계획, 검증 기능을 결합하는 기술이다. AI가 단순한 정보 생성 도구를 넘어 실제 업무와 물리시스템을 스스로 수행하고 제어하는 주체로 발전할수록, 뉴로심볼릭 AI의 중요성은 더욱 커질 것이다.

2. 뉴로심볼릭 에이전틱 AI의 기술과 기업 동향

뉴로심볼릭 AI는 단일 알고리즘이 아니라, 뉴럴네트워크와 심볼릭 시스템을 어떻게 결합하고 상호작용하는가에 따라 다양한 구조로 구현되는 기술 체계이다. 대표적으로 뉴럴네트워크가 인식한 정보를 기호로 변환한 후 논리적으로 추론하는 방식, 규칙과 지식을 학습 과정의 제약조건으로 반영하는 방식, 뉴럴네트워크와 심볼릭 시스템을 독립적으로 구성하고 반복적으로 정보를 교환하며 협력하는 방식 등이 연구되고 있다.[1]

심볼릭 시스템을 표현하는 대표적인 방식으로는 지식 그래프(Knowledge Graph) 계열과 규칙(Rule) 계열이 있다. 지식 그래프는 객체, 개념 및 이들 간의 관계를 구조화하여 표현하며, 규칙은 특정 조건에서 성립해야 하는 논리적 관계나 행동 조건, 제약조건을 명시적으로 표현한다. 이러한 심볼릭 시스템은 뉴럴네트워크만으로는 확보하기 어려운 지식 표현, 논리적 추론, 규칙 준수 및 결과 검증을 보완하는 역할을 한다.[2]

다만 지식 그래프와 규칙을 사람이 직접 구축하고 지속적으로 갱신하는 데에는 많은 비용과 전문지식이 요구된다. 이에 따라 최근에는 뉴럴네트워크를 이용해 데이터에서 객체, 관계, 규칙을 추출하고 이를 지식 그래프나 규칙 형태로 자동 또는 반자동 구축하는 방법도 함께 연구되고 있다. 즉, 뉴로심볼릭 AI는 뉴럴네트워크가 기호 지식을 활용하는 것뿐 아니라, 뉴럴네트워크를 통해 기호 지식 자체를 생성 및 보완하고 다시 추론에 활용하는 방식으로도 발전하고 있다.



이와 같이 뉴로심볼릭 AI는 단순히 뉴럴네트워크가 심볼릭 시스템을 이용하는 단방향 구조에서 벗어나고 있다. 뉴럴네트워크가 데이터로부터 새로운 지식과 규칙을 추출하고, 심볼릭 시스템이 이를 구조화하여 추론과 검증에 활용하며, 그 결과를 다시 뉴럴네트워크와 에이전트의 계획 및 행동에 반영하는 순환형 구조로 발전하고 있다.

특히 이러한 구조는 에이전틱 AI와 결합될 때 더욱 중요성이 커진다. 에이전트는 사용자의 목표를 이해하고 계획을 수립한 뒤 외부 도구와 시스템을 직접 실행하기 때문에, 단순히 올바른 답변을 생성하는 것보다 계획과 행동이 규정, 권한 및 안전조건을 만족하는지 검증하는 능력이 필요하다.

기업과 연구기관의 동향은 다음과 같다.

Google DeepMind는 AlphaGeometry와 AlphaProof를 통해 뉴럴네트워크 기반 탐색과 형식적 추론을 결합하는 연구를 발전시키고 있다. 특히 AlphaGeometry 2는 Gemini 기반 언어모델과 고속 기호 추론 엔진을 결합한 뉴로심볼릭 시스템으로, 뉴럴네트워크가 유망한 해법을 탐색하고 심볼릭 시스템이 논리적으로 검증한다. AlphaProof는 언어모델과 강화학습을 활용해 Lean 형식 언어에서 증명을 탐색하고 검증하는 구조를 사용한다. 두 시스템은 2024년 국제수학올림피아드 문제 6개 중 4개를 해결하여 은메달 수준의 성능을 보였다.[3][4]

AWS는 Amazon Bedrock Guardrails의 Automated Reasoning Checks를 통해 자연어로 작성된 정책과 업무 규칙을 심볼릭 시스템의 정책으로 변환하고, 파운데이션 모델의 응답이 정의된 규칙과 일치하는지를 검증하는 기능을 제품화하였다. 이는 기존의 키워드나 확률 기반 필터링과 달리 규칙과 변수에 기반하여 결과의 타당성을 검증하고 근거를 제시한다.[5]

NVIDIA는 GPU를 이용한 대규모 기호 추론의 가속 가능성을 제시하고 있다. RAPIDS와 Prometheux Vadalog Parallel은 지식 그래프와 도메인 규칙을 결합하여 기호 추론을 수행하는 시스템으로, Spark RAPIDS를 통합해 GPU를 통하여 연산을 가속화하였다. 생명과학 지식 그래프에서는 최대 9배, 기업 소유관계 그래프에서는 최대 3배의 성능 향상이 보고되었다. 이는 뉴로심볼릭 AI가 알고리즘뿐 아니라 GPU, 메모리, 분산 실행환경과 같은 컴퓨팅 및 시스템 SW 문제로 확장되고 있음을 보여준다.[6]

Microsoft는 LLM의 결과를 기호적 제약조건과 형식적 추론을 통해 검증하는 검증 가능 AI 연구를 강화하고 있다. 2026년 발표한 NSVIF는 자연어 지시를 논리적 제약과 의미적 제약으로 표현하고, 통합 증명기를 이용해 LLM의 출력이 해당 지시를 준수하는지 확인한다. 특히 에이전트의 다단계 실행 과정에서 작은 오류가 누적되는 문제를 대상으로 뉴로심볼릭 검증을 적용하고 있다.[7]

DARPA는 ANSR(Assured Neuro-Symbolic Learning and Reasoning) 프로그램을 통해 데이터 기반 학습과 기호 추론을 통합하는 고신뢰 AI 기술을 개발하고 있다. ANSR는 뉴로심볼릭 AI의 목표를 단순한 추론 성능 향상에 두지 않고, 공격에 대한 강건성, 안전성 판단을 위한 증거 확보, 주어진 명세에 대한 예측 가능성을 함께 확보하는 데 두고 있다.[8]

LinkedIn은 뉴럴네트워크를 이용해 지식 그래프를 자동으로 보완하는 사례를 보여주었다. LinkedIn의 Entity-BERT는 회원 프로필의 기존 엔티티 관계를 Graph Neural Network와 Transformer로 학습하고, 명시적으로 입력되지 않은 직무, 기술, 기업 등의 엔티티를 추론하여 회원 지식 그래프를 보완한다. 이는 사람이 모든 기호 지식을 직접 구축하는 방식에서 벗어나 뉴럴네트워크를 이용하여 기호 지식을 자동 생성하고 확장하는 접근의 대표적인 산업 사례이다.[9]

이러한 동향을 종합하면, 현재의 뉴로심볼릭 AI 기술은 지식 구축, 기호 추론, 형식 검증, 실행 통제, 컴퓨팅 가속 등 개별 영역을 중심으로 발전하고 있다. Google DeepMind는 추론과 증명, AWS와 Microsoft는 규칙 기반 검증, NVIDIA 및 관련 생태계는 대규모 기호 추론의 컴퓨팅 가속, LinkedIn은 뉴럴네트워크를 활용한 지식 그래프 구축이라는 서로 다른 영역에 강점을 보인다.

그러나 파운데이션 모델, 지식 그래프, 기호 추론, 에이전트 실행환경, 검증 및 가드레일, 이기종 컴퓨팅을 하나의 구조로 통합한 범용적인 '뉴로심볼릭 에이전틱 AI 컴퓨팅 플랫폼'은 아직 뚜렷하게 정립되지 않은 단계로 볼 수 있다. 특히 현재 상용화는 LLM 응답 검증이나 가드레일과 같은 일부 계층에서 먼저 진행되고 있으며, 이를 AI OS와 컴파일러, 이기종 컴퓨팅까지 연결하는 통합 시스템 기술은 상대적으로 초기 단계에 있다.

3. 신뢰할 수 있는 에이전트를 위한 뉴로심볼릭 AI 컴퓨팅 플랫폼

새로운 알고리즘을 개발하는 것 만으로는 산업 분야에 적용할 수 없다. 컴퓨팅 구조, 시스템 SW, AI OS, 안전 실행 통제, 개발 환경을 통합한 하나의 컴퓨팅 플랫폼으로 접근해야 한다. AI 컴퓨팅 플랫폼은 아래 그림과 같이 다섯 가지 기술 영역으로 구성된다.

[그림 3] 뉴로심볼릭 AI 컴퓨팅 기술 분류



첫째, 뉴로심볼릭 AI 엔진 및 지식체계

뉴럴네트워크와 심볼릭 시스템을 연결하여 학습, 추론, 계획을 수행하는 기술이다. 파운데이션 모델은 비정형 데이터를 인식하고 후보를 생성하며, 기호 추론기는 지식과 규칙을 기반으로 결과를 추론하고 검증한다. 이를 위해 뉴럴네트워크의 텐서와 임베딩을 심볼릭 시스템의 엔티티, 관계, 규칙으로 변환하고 다시 연결하는 인터페이스가 필요하다.

지식체계에서는 지식 그래프와 온톨로지가 개념과 관계를 구조화하고, 규칙과 제약조건이 허용 가능한 행동과 금지해야 할 행동을 정의한다. 특히 지식 그래프는 단순한 RAG(Retrieval-Augmented Generation, 검색 증강 생성)용 정보 저장소를 넘어 규칙, 제약조건, 권한 등을 포함하는 실행 가능한 지식체계로 발전시킬 필요가 있다. 이러한 지식을 학습, 추론, 계획, 기억에 활용하는 기능까지 통합한 뉴로심볼릭 AI 엔진이 핵심 산출물이다.

둘째, 이기종 컴퓨팅 및 메모리 시스템

뉴로심볼릭 AI의 다양한 연산을 효율적으로 처리하고 프로세서 간 데이터 교환을 지원하는 기술이다. 뉴로심볼릭 AI는 대규모 텐서 연산뿐 아니라 그래프 탐색, 규칙 매칭, 변수 바인딩, 제약조건 풀이와 같이 분기와 불규칙한 메모리 접근이 많은 연산을 함께 수행한다.

따라서 NPU와 GPU뿐 아니라 CPU, 필요에 따라 기호 추론 가속기를 결합하는 이기종 컴퓨팅 구조가 필요하다. 특히 인식, 추론, 검증, 행동이 반복되는 에이전트와 피지컬 AI에서는 정해진 시간 안에 검증과 실행을 완료하는 것을 보장하기 위해서는 새로운 컴퓨팅 구조가 필요하다.

또한 동일한 정보를 텐서, 벡터, 그래프, 기호 등 서로 다른 형태로 사용하는 만큼, 데이터 간 일관성을 유지하고 이동 비용을 줄일 수 있는 통합 메모리 구조가 필요하다. 다수 에이전트의 독립적인 실행과 보안을 지원하기 위해 메모리 보호, 자원 격리, 신뢰 실행환경 등 하드웨어 수준의 보호 기능도 요구된다.

셋째, 뉴로심볼릭 시스템 SW 및 AI OS

뉴로심볼릭 연산의 실행과 자원 배분을 관리하고, 다수 에이전트의 독립적인 실행과 협업을 지원하는 기술이다. 뉴로심볼릭 컴파일러와 런타임은 신경망 그래프뿐 아니라 기호 연산, 규칙, 제약조건, 에이전트 실행 흐름까지 분석하고, 이를 적절한 컴퓨팅 자원에 배치한다. 성능과 전력뿐만 아니라 실시간 제약과 검증 시간을 함께 고려한 실행 최적화가 필요하다.

AI OS는 에이전트별 실행환경을 제공하고 모델, 메모리, 도구, 데이터와 컴퓨팅 자원을 통합 관리한다. 에이전트 간의 역할을 조율(오케스트레이션)하고, 에이전트의 생명주기, 상태, 장기 기억, 사용 모델, 도구와 자원을 일관되게 관리하는 기능이 요구된다. 핵심 산출물은 뉴로심볼릭 컴파일러, 런타임, AI OS를 포함하는 통합 시스템 SW 플랫폼이다.

넷째, 신뢰성 검증 및 안전 실행 통제

AI가 생성한 계획과 행동을 검증하고, 위험 수준에 따라 실행을 허용하거나 제한하는 기술로 뉴로심볼릭 플랫폼의 핵심적인 차별 요소이다. 지식 그래프와 규칙 기반 명세를 이용하여 에이전트가 생성한 계획을 실행 전에 검증하고, 결과를 조건 충족, 조건 위반, 판단 불가로 구분할 필요가 있다. 조건을 만족한 경우에만 실행을 허용하고, 위반하거나 판단할 수 없는 경우에는 계획을 수정하거나 실행을 중지하고 필요하면 사람의 승인을 요구한다.

실행 중에도 에이전트가 여러 단계의 작업을 연속적으로 수행하므로 각 행동과 도구 호출을 실행 직전에 다시 확인해야 한다. 특히 정책 엔진과 검증기는 에이전트의 판단과 독립된 영역에 두어 에이전트의 오류나 공격이 검증 기능까지 함께 무력화하지 않도록 해야 한다.

에이전트 간 작업을 위임할 때에도 권한 확대를 방지해야 한다. 위임받은 에이전트는 위임한 에이전트에게 허용된 범위를 넘어서는 데이터, 도구, 자원에 접근할 수 없도록 정책적으로 제한해야 한다. 핵심 산출물은 규칙 기반 검증, 실행 가드레일, 보안 검증 및 안전 전환 기능을 통합한 검증 및 실행 통제 플랫폼이다.

다섯째, 개발운영 및 통합 평가

뉴로심볼릭 AI의 개발과 기존 시스템 전환, 운영 상태 관측, 변경 관리와 성능 평가를 지원하는 영역이다.

개발도구는 모델, 지식 그래프, 규칙, 기호 추론기, 시스템 SW와 연계되는 에이전트와 도구를 쉽게 개발할 수 있도록 지원해야 한다. 또한 기존 소프트웨어와 기존 AI 모델을 뉴로심볼릭 플랫폼의 에이전트와 도구 형태로 전환하는 기능도 필요하다.

특히 규칙과 지식은 모델과 마찬가지로 지속적으로 변경되므로 생성, 검토, 승인, 배포, 갱신의 관리체계를 갖추어야 한다. 각각의 규칙에는 출처, 적용 범위, 승인자, 버전과 변경 이력을 기록하여 추적 가능성을 확보해야 한다.

운영 단계에서는 사용 중인 모델, 지식 그래프, 규칙, 런타임의 버전 조합을 함께 기록하고, 규칙 위반, 판단 불가, 실행 실패 등의 발생 빈도를 지속적으로 관측해야 한다. 문제가 발생하면 이전 버전으로 되돌리거나 안전 모드로 전환할 수 있어야 한다. 핵심 산출물은 개발도구, AgentOps, 실행 추적 및 통합 평가 체계이다.



결론적으로, 뉴로심볼릭 AI는 특정 알고리즘 하나로 확보되는 것이 아니다. AI 엔진, 이기종 컴퓨팅, 시스템 SW, 검증 및 실행 통제, 개발운영 체계가 하나의 플랫폼으로 통합되어 함께 동작할 때 확보되는 시스템 수준으로 확보되어야 한다.

4. 향후 발전 방향과 시사점

뉴로심볼릭 AI는 과거의 심볼릭 AI로 돌아가는 기술이 아니라, 딥러닝의 인식과 생성 능력에 지식, 논리, 계획, 검증 능력을 결합하는 새로운 AI와 이를 위한 컴퓨팅 구조이다. 과거 심볼릭 AI가 한계를 보였던 지식 획득과 유지 문제도 LLM을 이용한 지식 생성과 함께 규칙의 검증, 승인, 갱신 및 이력 관리 체계를 통해 보완할 필요가 있다. 향후 발전 방향은 크게 세 가지로 볼 수 있다.

첫째, AI 경쟁의 중심이 모델에서 시스템으로 이동할 것이다. 에이전틱 AI와 피지컬 AI에서는 모델의 성능뿐 아니라 어떤 지식과 도구를 사용하고, 어떤 컴퓨팅 자원에서 실행하며, 오류가 발생했을 때 어떻게 검증하고 복구하는지가 중요하다. 따라서 파운데이션 모델, 지식체계, 추론기, 런타임, 컴퓨팅 자원 및 검증 기술을 하나의 시스템으로 통합하는 역량이 핵심 경쟁력이 될 것이다. AWS의 Automated Reasoning Checks[5]와 NVIDIA의 NeMo Guardrails[6]처럼 모델 외부에서 검증과 통제를 수행하는 기술이 제품화되고 있는 것도 이러한 변화를 보여준다.

둘째, AI 안전의 중심이 출력 검열에서 실행 검증으로 이동할 것이다. AI가 정보를 생성하는 데 그치지 않고 API 호출, 코드 실행, 금융 거래, 장비 제어와 같이 중요한 실제 행동을 수행하게 되면서, 무엇을 말했는가보다 무엇을 실행하는가가 더욱 중요해진다. 따라서 에이전트가 생성한 계획을 실행 전에 규칙, 권한 및 안전조건에 따라 검증하고, 실행 중에도 각 행동을 지속적으로 감시하는 기술이 필요하다.

셋째, 지식, 실행환경, 컴퓨팅을 연결하는 새로운 시스템 SW 계층이 중요해질 것이다. 생성형 AI가 대규모 텐서 연산을 중심으로 발전했다면, 뉴로심볼릭 에이전틱 AI는 그래프 탐색, 기호 추론, 계획, 검증 및 도구 실행과 같은 다양한 연산을 함께 수행한다. 이에 따라 CPU, GPU, NPU와 필요시 기호 추론 가속기를 결합하는 이기종 컴퓨팅과 이를 효율적으로 관리하는 컴파일러 및 런타임의 중요성이 커질 것이다. AI OS 역시 기존 Linux나 실시간 운영체제(RTOS)를 대체하기보다는 그 위에서 에이전트의 실행, 상태, 기억, 자원 및 권한을 관리하는 새로운 실행 계층으로 발전할 가능성이 높다. 지식 그래프도 단순한 검색용 지식 저장소에서 벗어나 객체 간 관계, 업무규정, 권한, 절차 및 물리법칙을 표현하고 에이전트의 계획과 행동을 검증하는 실행 가능한 지식체계로 발전해야 한다. 이때 중요한 것은 얼마나 많은 지식을 구축했는가보다, 누가 지식을 만들었고 어떤 근거를 가지며, 어떻게 검증되고 변경되었는지를 관리하는 것이다.

이러한 변화는 국내 AI 전략에도 중요한 시사점을 제공한다. 그동안 국내 AI 투자는 파운데이션 모델, 데이터, GPU와 NPU 등 모델과 컴퓨팅 인프라 확보에 상대적으로 집중되어 왔다. 그러나 모델과 반도체 사이에는 컴파일러, 런타임, AI OS, 지식체계, 검증 및 운영도구와 같은 시스템 SW 계층이 존재한다. 향후 에이전틱 AI와 피지컬 AI가 확산될수록 이러한 실행 계층의 중요성은 더욱 커질 것으로 예상된다.

특히 글로벌 기업들이 검증, 가드레일, 지식 관리 및 에이전트 실행환경을 빠르게 제품화하고 있다는 점을 고려하면, 국내에서도 AI의 실행과 검증을 담당하는 시스템 SW 기술을 전략적으로 확보할 필요가 있다. 그렇지 않을 경우 향후 AI 서비스뿐 아니라 에이전트와 피지컬 AI의 핵심 실행환경을 해외 플랫폼에 의존할 가능성이 커질 수 있다.

따라서 '신뢰성 있는 뉴로심볼릭 AI 컴퓨팅 플랫폼'을 전략 기술영역으로 설정하고, 파운데이션 모델, 지식 그래프, 기호 추론, 이기종 컴퓨팅, AI OS 및 실행 검증 기술을 통합적으로 개발할 필요가 있다. 또한 기술 개발에 그치지 않고, AI의 잘못된 판단과 실행이 큰 피해로 이어질 수 있는 국방, 자동차, 산업설비, 에너지 등 사회 인프라의 AX를 위해서 뉴로심볼릭 에이전틱 AI 컴퓨팅을 적용할 필요가 있다.



참고문헌

[1] Wang, W., Yang, Y., Wu, F. (2022). Towards Data- and Knowledge-Driven Artificial Intelligence: A Survey on Neuro-Symbolic Computing. arXiv preprint arXiv:2210.15889. https://arxiv.org/abs/2210.15889

[2] Sheth, A., Roy, K., Gaur, M. (2023). Neurosymbolic AI - Why, What, and How. arXiv preprint arXiv:2305.00813. https://arxiv.org/abs/2305.00813

[3] Chervonyi, Y., Trinh, T. H., et al. (2025). Gold-medalist Performance in Solving Olympiad Geometry with AlphaGeometry2. arXiv preprint arXiv:2502.03544. https://arxiv.org/abs/2502.03544

[4] Tsoukalas, G., et al. (2026). Advancing Mathematics Research with AI-Driven Formal Proof Search. arXiv preprint arXiv:2605.22763. https://arxiv.org/abs/2605.22763

[5] Amazon Web Services (2025). Automated Reasoning checks in Amazon Bedrock Guardrails. AWS What's New, August 2025. https://aws.amazon.com/about-aws/whats-new/2025/08/automated-reasoning-checks-amazon-bedrock-guardrails/

[6] NVIDIA. NeMo Guardrails Library - Rail Types. NVIDIA NeMo Documentation. https://docs.nvidia.com/nemo/guardrails/about-nemo-guardrails-library/rail-types

[7] Microsoft Research. Neuro-Symbolic Verification on Instruction Following of LLMs. Microsoft Research Publication. https://www.microsoft.com/en-us/research/publication/neuro-symbolic-verification-on-instruction-following-of-llms/

[8] DARPA. Assured Neuro-Symbolic Learning and Reasoning (ANSR). DARPA Program Page. https://www.darpa.mil/research/programs/assured-neuro-symbolic-learning-and-reasoning

[9] LinkedIn Engineering (2021). Completing a Member Knowledge Graph with Graph Neural Networks. LinkedIn Engineering Blog. https://engineering.linkedin.com/blog/2021/completing-a-member-knowledge-graph-with-graph-neural-networks