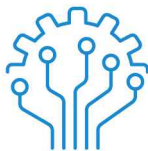
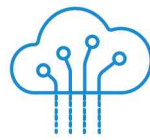


# AGI 기술개발 동향

## - 주요 빅테크 기업 중심으로



정선화



본 보고서는 ETRI ICT전략연구소 기본사업인  
"국가 지능화 기술정책 및 표준화 연구"를 통해 작성된 결과물입니다.



본 보고서의 내용은 연구자의 견해이며 ETRI의 공식 의견이 아님을 알려드립니다.

# 목 차 C O N T E N T S

## Executive Summary

1. 분석 배경 .....	1
2. AGI 정의 .....	2
3. AGI 구현을 위한 세부기술 .....	8
4. 빅테크 기업의 AGI 연구 및 개발 동향 .....	12
5. 국내 기업의 AGI 개발 동향 .....	32
6. 소결 .....	35
참고문헌 .....	37





## Executive Summary

## 📄 범용인공지능 (AGI, Artificial General Intelligence) 개요

- (정의) 다양한 작업에서 인간보다 뛰어난 능력을 보이는 인공지능으로 성과와 기능 측면에서 다소 모호하게 정의되던 AGI가 단일 시점의 획기적인 결과물이 아닌 다양한 능력의 점진적 통합과 실제 응용을 통한 발전 과정으로 새롭게 정의되고 있음
  - AGI를 개발하고 평가하기 위해서는 AGI가 무엇을 의미하는지 명확하고 구체적인 정의가 필요하고, 2024년에 AGI를 개발하는 빅테크 기업들 사이에서 본격적으로 논의 시작
  - (과거) 사람의 개입 없이 광범위하고 복잡한 작업을 수행할 수 있는 시스템으로 정의
  - (현재) 2024년 딥마인드, 오픈AI, JRI는 능력, 자율성, 활용범위 등을 고려하여 단계별로 발전해 나가는 기술로 AGI를 정의
    - \* 딥마인드는 올 초 능력과 활용범위 기준 5단계(신흥, 유능한, 전문가, 거장, 초인)로 정의('24.1)
    - \* 오픈AI 또한 7월에 딥마인드의 자율성 단계에 따른 AGI 정의와 유사하게 능력과 자율성을 기준으로 5단계 (도구, 조연자, 협력자, 전문가, 에이전트)로 정의 ('24.7)
- (중요성)
  - ChatGPT 공개 이후('22.11) 전문가들은 AGI 도래 시기를 기존보다 앞당겨 전망
  - 수년 이내 개발될 것이라는 전망 속에 빅테크기업들의 AGI 개발 경쟁은 가속화
  - 정치·경제·사회에서 파괴적 혁신을 가져올 AGI 기술패권을 위해 미중 경쟁 심화
    - \* 미 의회는 중국의 AI 기술발전에 대응하기 위해 대규모 자원 프로그램(AI 맨하튼 프로젝트)의 시급성 강조('24.11)
- (세부기술)
  - ① AGI 기본모델 : 사람의 뇌가 감각, 감정, 인지 영역으로 나뉘어 있는 것과 유사하게 AGI 기본모델에서 지각, 추론, 기억, 메타인지라는 4대 핵심 기능을 구현
  - ② 인터페이스 : 디지털 환경, 에이전트, 로봇과의 상호작용 능력 구현
  - ③ 시스템 효율화 : 확장 가능한 모델 아키텍처 설계, 대규모 학습 및 추론 최적화 기술 등
  - ④ 안전과 신뢰 : 안전성과 신뢰성 확보, 인간 가치를 정렬하기 위한 방법론 등
  - ⑤ 응용 : 과학을 위한 AI, 소프트웨어 자동 생성 기술, 다양한 사회 문제해결 분야 등

## 📄 주요 기업별 연구 및 개발 핵심 동향

- 대형언어모델을 기반으로 멀티모달리티를 강화하고 추론 성능 향상을 위해 노력 중
- 디지털기기들과 상호작용하는 에이전트 서비스 경쟁 심화

### ① 구글

- 現 대형언어모델의 주류 아키텍처인 트랜스포머(Transformer)를 개발한 구글은 언어모델 고도화, 멀티모달학습 강화, 추론능력 개선이라는 **소프트웨어 혁신**과 함께 AI 전용 칩과 양자 컴퓨팅이라는 **하드웨어 혁신**을 동시에 추구하면서 보다 효율적이고 범용적인 AGI 개발을 위한 통합적 접근을 시도 중

### ② 딥마인드

- 뛰어난 강화학습 능력을 기반으로 알파고, 알파제로, 알파폴드 등 특정 도메인에서 인간 수준을 능가하는 AGI를 개발한 딥마인드는 최근에 AGI 5단계 초인(Super Intelligence) 수준의 지능을 달성하기 위해서는 오픈엔드 학습 중요성을 강조하며 **오픈엔드 기초 모델 개발 방향**을 제시하고 있으며 특히 수학적 추론과 과학적 문제해결 문제에서 주목할 만한 성과를 도출

### ③ 오픈AI

- GPT 시리즈에서 시작하여 최근 출시된 o1 모델까지 언어모델의 성능을 획기적으로 발전시켰으며 멀티모달능력 강화, 플러그인을 통한 도구 활용성 확장, 콘텐츠 조정 자동화와 성능평가 시스템 구축 등 포괄적이고 실용적인 접근으로 AGI 개발을 추구하며, 자체 정의한 AGI 2단계의 추론자 수준을 달성하고 **에이전트 단계로 진입**

### ④ 메타

- 효율적인 언어모델 개발, 다중 모달리티의 통합, 메모리 시스템 혁신, 그리고 실세계 응용을 위한 기술 개발을 **오픈소스 전략**을 통해 협력적 발전을 추구하고, **대규모 인프라 투자**를 통해 AGI 개발에 필요한 컴퓨팅 능력도 확보

### ⑤ 앤스로픽

- **스케일링 가설**에 기반한 모델 성능 극대화를 추구하면서도, **AI 시스템의 안전성과 윤리성**을 핵심 가치로 삼아 책임 있는 AGI 개발을 위해 노력하고 있으며 모델의 해석 가능성 향상과 환각 현상 감소 등 AI 안전 관련 연구에서 성과 도출

## 1 분석 배경

- 2022년 11월 ChatGPT 공개 이후 전문가들은 AGI 도래 시기를 기존보다 **앞당겨** 전망
  - 2023년 수행된 ESPR 설문조사<sup>1)</sup>에서 전문가들은 HLMI(High-Level Machine Intelligence)가 **50% 확률**로 달성 예측되는 해를 **2047년**도로, **10% 확률**의 경우 **2027년**도로 전망
    - 전년도 조사에서는 동일 질문에 대해 50% 확률로 2060년, 10% 확률로 2029년을 전망 했으므로 결과적으로 1년 사이에 각각 13년, 2년 앞당겨 예측됨
  - AI 빅테크 기업 리더 및 개발자들은 **향후 5년 이내**로 AGI 개발 시기를 더 **전향적으로** 전망
    - 오픈AI 샘 알트먼(Sam Altman)은 ‘**가까운 미래**’에 달성될 것으로 전망<sup>2)</sup>
    - 딥마인드 셰인 레그(Shane Legg)는 50%의 확률로 AGI 달성시기를 **2028년**으로 전망<sup>3)</sup>
    - 테슬라 일론 머스크(Elon Musk)는 초인 수준의 AGI가 **2026년**에 개발될 것으로 전망<sup>4)</sup>
- 수년 이내로 AGI가 개발될 것이라는 전망 속에 **빅테크 기업들의 AGI 개발 경쟁은 가속화**
  - AI 선도 빅테크 기업들은 **AGI 개발을 최우선 과제로 선정**하고 기술패권 확보에 노력 중
    - 2023년 4월, 구글은 자체 AI 연구팀인 구글 브레인과 딥마인드를 통합하여 ‘구글 딥마인드’를 설립하고 AGI 개발에 인력과 자원을 결집함
    - 오픈AI는 설립 당시부터 AGI 개발에 초점을 맞추고 있었으며, 이와 관련된 효율적 의사결정을 위하여 영리 기업으로 전환을 모색 중이라는 기사가 최근 보도되기도 함<sup>5)</sup>
    - 메타는 라마 시리즈 모델 공개 이후 대규모언어모델의 오픈소스 생태계를 주도 중
  - 2023년 글로벌 경기침체에도 불구하고 **AI 분야 스타트업 투자는 양호한 회복력**을 보였으며, 미국의 빅테크 기업들에 **투자자금이 쏟아지는 양상**을 보임<sup>6)</sup>
    - 전체 벤처캐피털 시장은 2022년 대비 42% 감소하였으나, AI 분야는 10% 감소에 그쳐 상대적으로 양호한 투자회복력을 시현
    - 오픈AI, 앤스로픽, 인플렉션AI 등 대규모언어모델 개발 선도 기업들의 메가라운드 투자 유치에 힘입어, 미국의 투자 규모는 전년 대비 14% 증가한 반면, 유럽은 29% 감소하고, 아시아는 무려 61% 급감하며 뚜렷한 지역 편중 현상 발생

1) AI Impact에서 수행한 ESPR(Expert Survey on Progress in AI) 설문조사는 1,700여 명의 전문가 집단을 대상으로 2023년 10월에 수행되었으며, 이 중 HLMI에 대한 문항의 설문결과를 제시

2) 오픈 AI CEO 샘 알트먼의 CNBC와 인터뷰에서 가까운 미래(reasonably close-ish future)를 언급 (2024.1.16.) (출처: <https://www.cnbc.com/2024/01/16/openai-sam-altman-agi-coming-but-is-less-impactful-than-we-think.html>)

3) 딥마인드 공동 창업자 Shane Legg 인터뷰 (2023.11) (출처: <https://www.youtube.com/watch?v=Kc1atfjkijU>)

4) 로이터통신 보도 (2024.4.9.) (출처: <https://www.reuters.com/technology/teslas-musk-predicts-ai-will-be-smarter-than-smartest-human-next-year-2024-04-08/>)

5) AP통신 보도 (2024.9.27.) (출처: <https://apnews.com/article/chatgpt-openai-sam-altman-nonprofit-859bff5c19845f51796244e0072e2dfb>)

6) State of AI 2023 Report (2024.2.1.) (출처: <https://www.cbinsights.com/research/report/ai-trends-2023/>)

- 생성형 AI가 전체 AI 분야 투자의 48%를 차지하며 시장 주도권 확보
- \* 2022년 AI 분야 투자의 8% 비중에서 2023년 48%로 대폭 상승

## 2 AGI 정의

- (정의) 다양한 작업에서 인간보다 뛰어난 능력을 보이는 인공지능으로 성능과 기능 측면에서 다소 모호하게 정의되던 AGI가 단일 시점의 획기적인 결과물이 아닌 **다양한 능력의 점진적 통합과 실제 응용을 통한 발전 과정**으로 인식 전환 중
  - AGI를 개발하고 평가하기 위해서는 AGI가 무엇을 의미하는지 명확하고 구체적인 정의가 필요하고, 2024년부터 AGI를 개발하는 빅테크 기업들 사이에서 논의가 시작
  - (과거) 사람의 개입 없이 광범위하고 복잡한 작업을 수행할 수 있는 시스템으로 정의
    - 1956년 다트머스 회의에서 인공지능 개념이 발표되고 얼마 지나지 않아 1960년대부터 일반적인 지능이라는 개념이 등장<sup>7)</sup>
    - AGI이라는 용어의 최초 사용은 1997년 마크 구브루드(Mark Gubrud)의 글에서 시작 되었으며 활용 범위 이외에도 인간의 뇌에 필적하거나 능가하는 능력을 강조<sup>8)</sup>
    - \* 복잡성과 속도 면에서 사람의 뇌에 필적하거나 능가하는 수준으로 지식을 습득하고 조작하고 추론할 수 있으며 사람의 지능이 필요한 모든 산업 또는 군사 작전에서 기본적으로 사용할 수 있는 AI시스템으로 정의
    - 셰인 레그(Shane Legg)와 벤 고어첼(Ben Goertzel)은 2000년대 컴퓨터과학자들 사이에서 AGI라는 용어를 대중화
    - \* AGI는 사람이 일반적으로 수행할 수 있는 인지작업을 수행할 수 있는 기계로 설명
  - (현재) 2024년 딥마인드, 오픈AI, JRI는 **능력, 자율성, 활용범위** 등을 고려하여 **단계별로 발전해 나가는 AGI 정의를 발표**
    - 딥마인드는 **올 초 능력과 활용범위** 기준 5단계로 정의('24.1), **현 범용목적 1단계** 수준<sup>9)</sup>
    - \* AGI 정의에 다음의 6가지 원칙 적용
      1. 알고리즘보다는 실제 능력에 초점
      2. 범용성과 능력을 함께 고려
      3. 인지 및 메타인지 작업에 초점 - 행동은 고려하지 않음
      4. 배포 기준이 아닌 잠재적 성능에 초점
      5. 사람들이 중요하게 여기는 실제 작업에 초점
      6. AGI는 단일 시점의 결과물이 아니라 단계별 실현

7) 정치학자이자, 행정학자, 경제학자, 심리학자, 컴퓨터과학자인 허버트 사이먼(1978년 노벨경제학자, Herbert Simon)은 1965년에 20년 안에 인간이 할 수 있는 모든 일을 기계가 할 수 있을 것으로 예측

8) Mark Gubrud. Nanotechnology and International Security. Fifth Foresight Conference on Molecular Nanotechnology, November 1997.

9) Levels of AGI: Operationalizing Progress on the Path to AGI, Google DeepMind, arXiv:2311.02462v2 (2024.1)



[딥마인드의 AGI 정의 - 능력과 활용범위]

단계	이름	성능	특수목적	범용목적
1	신흥(Emerging)	숙련되지 않은 성인 수준	GoFAI 등	ChatGPT, 라마2 등
2	유능한(Competent)	숙련된 성인의 상위 50% 이상	시리, 알렉사 등	-
3	전문가(Expert)	숙련된 성인의 상위 10% 이상	그래머리, 달리2 등	-
4	거장(Vertuso)	숙련된 성인의 1% 이상	딥블루, 알파고	-
5	초인(Superhuman)	숙련된 성인 능력 초월	알파폴드, 알파제로 등	-

- 또한, 딥마인드는 작업에서의 자율성을 기준으로 인간-AGI 상호작용 패러다임 정의

[딥마인드의 AGI 정의 - 자율성]

단계	이름	자율성 수준	자율성과 AGI 단계 매칭	
			최소 단계	완성 단계
1	도구(Tool)	사람이 100% 통제	특수목적 1단계	특수목적 2단계
2	조언자(Consultant)	사람이 요청시만 작업	특수목적 2단계	특수목적 3단계 범용목적 1단계
3	협력자(Collaborator)	사람-AI 대등 협력 작업	범용목적 1단계	범용목적 2단계
4	전문가(Expert)	AGI 주도 상호작용	특수목적 4단계	범용목적 3단계
5	에이전트(Agent)	AGI 100% 자율적 작업	범용목적 4단계	범용목적 5단계

- 오픈AI 또한 7월에 딥마인드의 자율성 단계에 따른 AGI 정의와 유사하게 능력과 자율성을 기준으로 5단계로 정의<sup>10)</sup> ('24.7), 現 AI 수준 1단계 수준

[오픈AI의 AGI 정의]

단계	이름	설명
1	챗봇(Conversational AI)	현재 대중화된 AI 수준
2	추론가(Reasoning AI)	복잡한 문제를 해결 가능한 작은 문제로 분해하고, 작은 문제의 해결 순서와 통합적인 출력 결정
3	에이전트(Autonomous AI)	다면적인 문제를 탐색-해결하기 위해 반복적인 자체평가 및 자기개선 수행
4	혁신가(Innovating AI)	새로운 문제를 정의하고 해결하는 창의성 발현
5	조직(Organizations)	여러 워크플로우를 동시에 처리하고 문제해결 과정을 가속화

- 일본종합연구소(JRI)는 능력, 활용범위, 자율성 모두를 고려하여 3단계로 제시<sup>11)</sup> ('24.6) 하였으며, AGI 조건으로 신체성을 명시한 점이 딥마인드와 오픈AI와의 차별점

\* 구글 딥마인드와 오픈AI는 물리세계 이해가 AGI 실현에 필수적이라고 보지만, 학습데이터 부족과 시간·비용 문제로 지각 및 인지 영역 중심의 AGI 개발에 현재는 집중 중

10) OpenAI's 5 Levels of Super AI (2024.7) (출처: <https://www.forbes.com/sites/jodiecook/2024/07/16/openai-5-levels-of-super-ai-agi-to-outperform-human-capability/>)

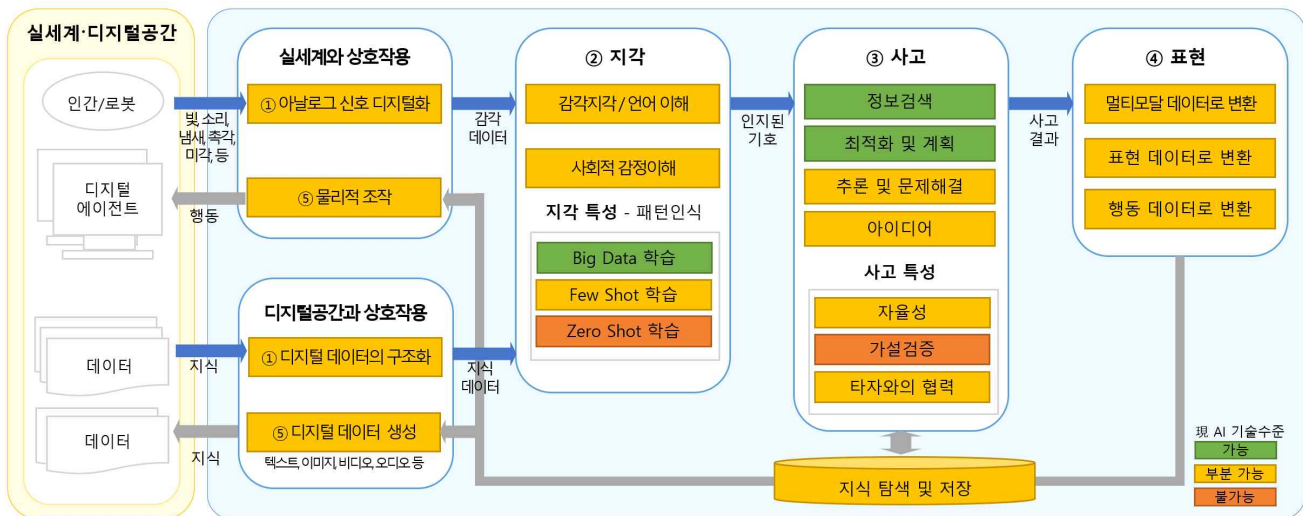
11) 汎用AIの現状と展望 - 実現への見通しと社会への影響, JRI (2024.6)

[일본종합연구소의 AGI 정의]

	2024~2026년	2027~2029년	2030년~
	자율성은 부족하지만 능력은 뛰어난 제한적 AI	디지털 공간에서 자율성을 확보한 범용 AI	실세계에서 자율성을 확보한 범용 AI
학습	- 범용 AI 기본능력 향상 - 학습 효율 개선	- 디지털 세계에서 자율성 실현 - 자율적 지식 탐색 - 자율학습 강화	- 신체성 기반 학습
외부와 상호작용	- GUI 조작 가능 - 다수 시스템 조작 가능	- 소프트웨어 자율 조작 - 디지털 공간에서 인간과 소통하며 일반 사무 가능	- 실세계에서 인간작업을 로봇으로 대체

- AGI 정의에 대해 명확한 합의에 도달하지는 못했지만 논의가 시작되었다는 점에서 의미
  - 단계별 정의 논의 및 합의 도출 과정은 現 기술수준을 평가하고 향후 기술로드맵을 수립하는 데 중요한 역할을 함으로써 AGI 실현을 위한 기술혁신을 이끄는 기반이 될 것임

- (기능) JRI와 맥킨지는 現 AI가 AGI로 발전하기 위해 반드시 확보해야 할 핵심 기능을 제시
- 일본종합연구소(JRI)는 다양한 작업을 사람과 동등한 수준으로 실현하기 위해 필요한 기능으로 지각인식, 계획 및 추론, 자율성장, 사회성, 신체성을 도출



[일본종합연구소의 AGI 기능 정의]

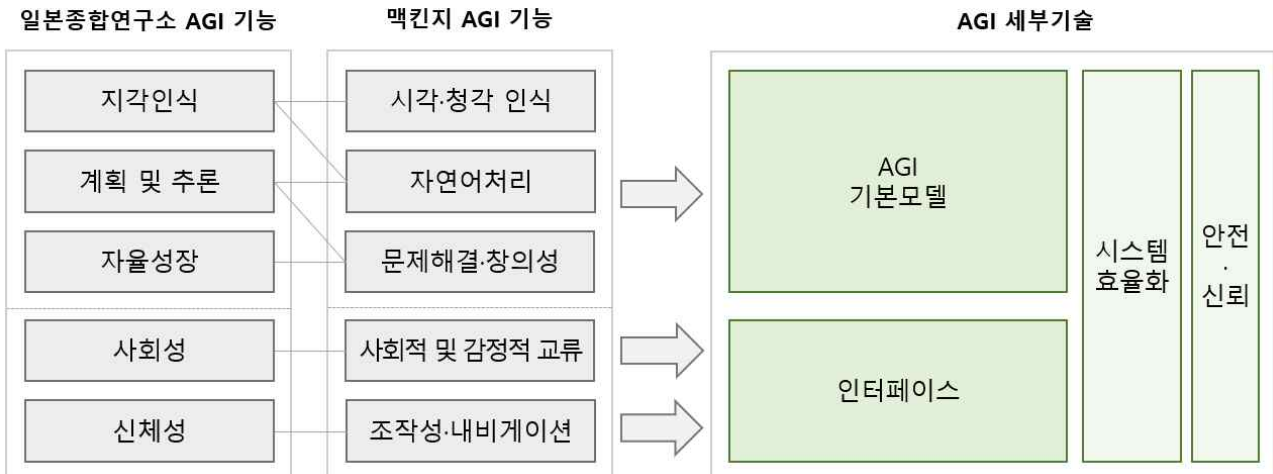
- 맥킨지는 AGI가 인간처럼 사고하고 학습하며 사람과 상호작용하기 위해 필요한 8가지 핵심 기능으로 시각·청각 인식, 자연어처리, 문제해결, 창의성, 사회성, 운동성을 제시<sup>12)</sup>

12) McKinsey & Company, What is artificial general intelligence (AGI)? (2024.3)

[맥킨지의 AGI 핵심 기능 제시]

기능	설명
시각인식 (Visual Perception)	인간 수준의 감각 인식을 달성해야 하며, 이를 통해 환경의 색상 일관성 등을 정확히 인식해야 함
청각인식 (Audio Perception)	소리의 공간적 특성을 파악할 수 있어야 하며, 이를 통해 소리의 위치나 방향을 정확히 인식할 수 있어야 함
자연어처리 (Natural Language Processing)	인간 수준의 일반 지식과 상식을 바탕으로 인간이 의사소통할 때 언급하지 않는 많은 정보를 이해하고 맥락을 파악할 수 있어야 함
문제해결 (Problem-Solving)	환경과 경험에서 학습하여 새로운 상황에 적응하고 문제를 인식하고 해결할 수 있어야 하며, 이를 위해 상식이나 시뮬레이션을 통해 가능성과 확률을 판단하는 능력 필요
창의성 (Creativity)	인간 수준의 창의성을 가져야 하며, 자신을 개선하기 위해 새로운 코드를 작성하거나 아이디어를 제시할 수 있어야 함
사회적 감정 교류 (Social and Emotional Engagement)	인간의 표정이나 음성 변화 등을 통해 감정을 해석하고, 인간과의 상호작용에서 공감을 보여야 함
정교한 조작 기술 (Fine Motor Skills)	섬세한 조작 기술을 가져야 하며, 이를 통해 머리카락을 뜯거나 수술과 같은 복잡한 작업을 수행할 수 있어야 함
내비게이션 (Navigation)	GPS와 같은 기술을 통해 자율적으로 환경을 탐색하고 상호작용할 수 있어야 함

- AGI 핵심 기능을 실현하기 위해서는 **세부기술들이 유기적으로 결합되고 상호연계되어야** 하며 인간과의 안전한 공존을 위해 **신뢰성 확보와 시스템 효율화 연구 또한 중요**
- 계획 및 추론 기능은 자연어처리와 문제해결·창의성 기능과 연결되며, 지각인식 기능은 시각·청각 인식 및 자연어처리 기능과 상호보완적으로 연결



[AGI 핵심 기능과 이를 실현하기 위한 세부기술 구성]

- 이처럼 핵심 기능들이 중복되거나 서로 연계됨에 따라 세부기술들도 상호보완적인 방식으로 개발될 수밖에 없음
  - AGI 개발에서 안전, 신뢰성, 효율성이 중요한 요소로 강조되며, 특히 고도화된 AGI가 가져올 수 있는 잠재적 위험을 대비하여, 안전성과 신뢰성을 강화하는 연구는 필수적
- (성능평가) 인간 수준의 다양한 인지 및 문제해결 능력을 갖춘 AGI를 평가하기 위해 새로운 벤치마크가 제안되고 있으며 AGI 기술 향상과 함께 지속적인 연구와 발전이 이루어지고 있음
- (평가항목) 인간 수준의 지각·사고·소통 능력과 윤리 및 안전성을 평가하는 항목으로 구성
    - (일반화) 학습한 데이터나 상황이 아닌 경우에도 일반화하고 적용할 수 있는 능력
    - (추론) 복잡한 문제를 분석하고 논리적으로 해결책을 도출하는 능력
    - (장기학습) 정보를 기억하고 새로운 정보와 통합하여 장기적으로 학습하는 능력
    - (의사소통) 언어를 이해하고 생성하며 의도를 파악하고 대화하는 능력
    - (안전) 윤리적 의사결정 능력과 인간의 통제 하에 안전하게 작동 가능한지 평가
    - 이러한 성능평가 연구는 AGI 성능이 향상되면서 안전하고 책임감 있게 개발될 수 있도록 방향을 제시
  - (벤치마크) 최근 3년 동안 많은 새로운 벤치마크가 제안되고 있으며 평가방법론도 빠르게 발전
    - AI Index 보고서에는 다양한 벤치마크를 사용하여 최신 AI 모델들을 평가하여 왔으나, 최근 AI 성능의 급격한 향상에 따라 기존의 벤치마크 성능이 포화되어 2024년 공개된 보고서에서는 코딩, 추론, 에이전트 행동 등을 평가하는 새로운 벤치마크 18개를 소개
      - \* 기존에 사용되었지만 2024 AI Index 보고서에서 사용되지 않는 벤치마크는 ImageNet, PubMed, SuperGlue, VQA(Visual Question Answering Challenge) 등
    - 새로운 벤치마크는 2021년부터 2023년으로 최근 3년 사이에 개발되었으며 주로 코딩, 추론, 에이전트 능력을 평가하는 벤치마크로 구성
    - 이 중 특정 도메인에서 성능을 평가하는 벤치마크를 제외하고, AGI 평가항목 기준으로 분류하고 개발기관, 출시년도, 주요 특징을 파악
      - \* 코딩(HumanEval, SWE-bench), 이미지생성(HEIM), 이미지편집(EditVal)과 관련된 4개의 벤치마크 제외
    - 이 외에 오픈AI, 딥마인드, 구글 등의 협업으로 인간 수준의 지능을 평가하기 위해서 개발된 BIG-Bench(Beyond the Imitation Game Benchmark)<sup>13)</sup>, GPT-4 모델을 평가하기 위해서 사용된 AI2에서 개발한 ARC(AI2 Reasoning Challenge)<sup>14)</sup> 등도 존재
    - 또한 뉴욕대에서 개발한 질문응답 모델에서 인종, 성별, 종교 등 다양한 요소의 편향을

13) Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models, arXiv:2206.04615 (2022.9)

14) Think you have Solved Question Answering? Try ARC, the AI2 Reasoning Challenge, arXiv:1803.05457, (2018.4)

측정할 수 있는 BBQ(Bias Benchmark for QA)<sup>15)</sup>, 유럽(튀빙겐대, EPFL)과 미국(퍼듀대, 프린스턴대)이 협업하여 개발한 악의적 공격에 대한 취약성을 평가하는 RobustBench<sup>16)</sup>, 강화학습 에이전트의 안전성을 평가하기 위해 Partnership on AI에서 개발한 SafeLife<sup>17)</sup> 등 윤리 및 사회적 영향력을 평가하는 다양한 벤치마크에 대한 연구가 활발하게 이루어지고 있음

- AGI를 평가하기 위한 주요 기준들은 서로 독립적인 단일 평가로 정확히 측정하기 어렵고, 하나의 벤치마크에서 이러한 기준들이 중첩되거나 혼합되어 복합적으로 측정됨

[AGI 평가를 위한 주요 벤치마크]

평가항목	이름	개발기관	출시년도	주요 특징
일반화	GPQA	뉴욕대학, 코히어, 앤스로픽	2023	다양한 분야의 문제해결 능력, 논리적 추론
	MMML	워털루대학, 오하이오대학, CMU, 등	2023	전문가 수준의 AGI 성능 평가
	BigTom	스탠포드대학	2023	복잡한 시나리오에서의 원인-결과 분석
추론	GSM8K	오픈AI	2021	수학적 문제해결 능력 평가
	MATH	UC버클리	2021	고급 수학 문제해결 능력 측정
장기 학습	HELM	스탠포드대학	2021	언어모델의 종합적 평가
	HaluEval	렌민대학, 몬트리올대학	2023	응답의 사실성, 정확성, 일관성
	TruthfulQA	옥스퍼드대학, 오픈AI	2021	정보의 정확한 기억과 활용 능력 평가
의사 소통	Chabot Arena Leaderboard	LMSYS Org	2023	인간 평가자들의 블라인드 테스트를 통한 상대적 순위 산정
	VisIT-Bench	구글, UCLA, AI2, 스탠포드대학 등	2023	이미지 이해, 지시사항 준수, 작업 수행 정확도
안전	MoCa	스탠포드대학	2023	윤리적 딜레마 해결, 도덕적 판단
에이전트	AgentBench	칭화대학, 오하이오대학, UC버클리	2023	웹 브라우징, API 사용, 데이터베이스 조작 등의 실제 태스크 수행 능력
	MAgentBench	스탠포드대학	2023	모델 학습, 디버깅, 최적화 능력
	PlanBench	아리조나대학, 콜로라도대학	2023	행동 및 변화에 대한 계획 또는 추론 능력

- o (한계와 과제) AGI 기술 발전과 함께 추론이나 상식적 이해를 평가하는 수준도 변화할 수 밖에 없으며, 포괄성, 객관성, 확장성을 고려한 지속적인 평가방법 연구 필요

15) BBQ: A Hand-Built Bias Benchmark for Question Answering, arXiv:2110.08193 (2021.10)

16) RobustBench: a standardized adversarial robustness benchmark, arXiv:2010.09670 (2020.10)

17) <https://github.com/PartnershipOnAI/safelife>

- 추론성능 향상 경쟁이 본격화되면서 에포크AI는 고차원적 추론과 창의력을 요구하는 고난이도 수학문제로 구성된 FrontierMath<sup>18)</sup>라는 새로운 벤치마크를 공개
- \* 오픈AI의 GPT-4o와 앤스로픽의 Claude도 벤치마크 문제 중 2%정도 밖에 풀지 못한 고난이도 문제로 구성
- 최근 에이전트로 AI기술 활용이 증대됨에 따라 오픈AI는 자율 머신러닝 엔지니어링 작업을 평가하기 위한 MLE-bench<sup>19)</sup>를 개발하고 자사 최신 모델 o1을 테스트
- ETRI는 최근 로봇에서 활용되는 AI 기술의 평가를 위해서 LoTa-Bench<sup>20)</sup>를 개발
- AGI 발전을 평가하기 위한 벤치마크는 이처럼 고정된 것이 아니며, AI기술의 발전속도를 고려하여 빠르게 대응하며 새로운 기능 및 성능을 평가할 수 있는 연구개발 필요
- 또한 AGI 벤치마크는 단일 벤치마크로 모든 능력을 평가하기 어렵기 때문에 다양한 능력을 균형 있게 측정하기 위한 다양한 벤치마크가 필요하며, 벤치마크에는 평가기준의 표준화, 문화적 편향 제거를 위한 노력도 함께 이루어져야 함

### 3 AGI 구현을 위한 세부 기술

\* 일본종합연구소와 맥킨지가 정의한 AGI 핵심기능과 이를 실현하기 위한 세부기술을 2024년 ICLR 논문<sup>21)</sup>을 참고하여 정리

- (AGI 기본모델) 사람의 뇌가 감각, 감정, 인지 영역으로 나뉘어 있는 것과 유사하게 AGI 파운데이션모델에서 지각, 추론, 기억, 메타인지라는 4대 핵심 기능을 구현
  - (지각) 시각, 청각 등 다중 감각 데이터를 통합적으로 처리하고 이해하는 능력
    - 초기에는 자연어 처리에 특화된 대규모언어모델(LLM, Large Language Model)이 주를 이루었으나, 최근에는 이미지, 비디오, 오디오 등 다양한 형태의 데이터를 통합적으로 처리하고 이해하는 멀티모달모델로 발전 중
    - 향후에는 다중 모달 데이터의 동시 처리 능력을 더욱 고도화하는 한편 각 모달리티간 상호작용과 모델 내부의 작동 원리를 이해하고 설명할 수 있는 수준의 해석 가능성 확보 연구가 필요
  - (추론) 광범위하고 복잡한 문제를 해결하기 위한 인간 수준 또는 이상의 추론 능력
    - 대규모언어모델은 사전학습된 지식을 바탕으로 추가적인 학습 없이도 과제수행이 가능한 제로샷(zero shot) 학습과 소수의 예시만으로 새로운 과제에 적응할 수 있는 퓨샷(few shot) 학습 능력을 보유하여 범용적인 문제해결과 일반화된 지능을 일부 구현

18) FrontierMath (출처: <https://epoch.ai/frontiermath>)

19) MLE-bench (2024.10.10.) (출처: <https://openai.com/index/mle-bench/>)

20) LoTa-Bench: Benchmarking Language-oriented Task Planners for Embodied Agents, arXiv:2402.08178 (2024.2.13.)

21) How Far Are We From AGI?, arXiv:2405.10313v1 (2025.5.16.)



- 현재 CoT(Chain of Thought), ToT(Tree of Thoughts) 등 다양한 추론 메커니즘을 활용하여 복잡한 문제를 단계적으로 분해하고 해결하며, 동적 계획 수립과 자기일관성을 유지하는 고도화된 추론 능력 연구 중
- 향후에는 이해와 일반화 능력을 향상시키기 위한 인과관계 학습 및 다단계로 이루어진 복잡한 과제를 자율적으로 해결하고 긴 맥락에서도 일관된 논리를 유지할 수 있는 긴 맥락 추론(long context reasoning) 연구가 필요
- 또한 마음이론(Theory of Mind) 기반의 타인이해 능력과 사회적 상호작용 능력을 강화하고 의사결정 과정의 투명성과 설명 가능성을 확보하는 연구도 필요
- \* 마음이론은 다른 사람의 정신상태, 즉 신념, 의도, 욕구, 감정 등을 이해하고 추론하는 능력
- (기억) 단기 및 장기기억을 효율적으로 저장하고 검색하는 능력
  - 에이전트가 현재 상황 정보를 짧은 시간 동안 유지하고 즉각적으로 활용할 수 있는 방법을 탐구하는 단기기억 연구와 에이전트의 과거 경험과 축적된 지식을 저장, 갱신, 활용 방법을 탐구하는 장기기억 연구로 구성
  - \* 에이전트는 사람의 개입 없이 특정 작업을 수행하는 자율지능형 시스템
  - 단기기억 관련하여 파운데이션 모델에서 상황 내 학습(in-context prompting), 실시간 데이터와 추론 결과의 즉시적인 활용 등 연구
  - 장기기억 관련하여 기억의 체계적인 구조화, 외부 데이터베이스나 웹 상의 정보 활용, 기억 검색 알고리즘 개선, 상황에 맞는 기억을 자동으로 추출하는 기법 등 연구
- (메타인지) 에이전트가 자신의 지식상태, 학습 및 사고 과정을 스스로 평가하는 능력
  - AGI가 스스로 지식과 능력에 대한 확신을 판단하는 자기평가모델, 작업 중 발생하는 오류를 자체 감지하고 필요시 수정하는 메카니즘, 메타인지와 강화학습을 결합하여 과거 경험을 통해 스스로 학습전략을 개선하는 등 다양한 연구가 진행 중
  - 향후에는 메타인지의 다양한 기능을 통합하여 보다 정교한 자율학습과 문제해결 능력을 향상시킬 수 있는 연구가 필요
- (인터페이스) 디지털환경, 에이전트, 로봇과의 상호작용 능력 구현
  - (디지털환경) 인터넷, 데이터베이스, 코드, API 등 디지털환경과 상호작용하며 사람과 유사한 지능적 행동을 수행하는 인터페이스
    - 외부 도구의 자율적 학습 및 활용, 다양한 API 연결 및 활용, 도구 생성 및 수정 등 대형언어모델과 디지털 도구를 통합 중
    - 향후에는 사람이 설계한 프레임워크를 넘어 자율적이고 독창적인 도구 개발 등 도구 활용 능력을 고도화하는 연구뿐만 아니라, 웨어러블 컴퓨팅, 스마트 환경, 혼합현실 등으로 디지털환경 확장 연구 등 필요

- (에이전트) AGI 에이전트 사이 상호작용을 위한 인터페이스
  - 강력한 모델이 약한 모델을 지도하는 지식증류(Knowledge Distillation) 방식과 다수의 AI가 동시에 협력하여 문제를 해결하는 방식을 연구 중
  - \* 지식증류란 더 큰 모델의 지식을 더 작은 모델로 전달하는 기술로, 성능이 뛰어난 대형 모델의 출력을 학습 데이터로 활용하여 더 작고 효율적인 모델을 훈련시키는 방법
  - 향후에는 효과적인 에이전트 네트워크 구축, 안전하고 윤리적인 상호작용 프로토콜 개발 등 필요

- (로봇) 사람이 로봇과 상호작용하기 위한 인터페이스
  - 자연어 명령을 로봇동작으로 변환하고, LM-Nav(Language Model for Navigation) 기술로 내비게이션을 수행하며 사람 동작을 이해하고 복제하는 수준에 도달
  - \* LM-Nav는 언어, 시각, 행동을 위한 사전학습 모델들을 결합하여 로봇 내비게이션을 수행하는 기술로 자연어 명령을 입력받고, 시각정보로 환경을 인식하고, 이를 바탕으로 로봇의 이동 경로와 행동을 계획하고 실시간으로 경로를 조정하며 목표지점까지 이동하는 기술
  - 실제 환경 기반 데이터셋을 통해 로봇의 일반화 능력이 향상되고 있으며, 향후에는 다중 감각 통합, 저렴한 로봇 제조 기술 발전, 엣지 컴퓨팅의 효율성 개선 등 연구 필요

□ (시스템효율화) 확장 가능한 모델 아키텍처 설계, 대규모 학습 및 추론 최적화 기술 등

- 확장 가능한 모델 아키텍처 설계
  - 셀프 어텐션 패턴의 최적화를 통한 계산 효율성 향상, 지식증류, 모델 프루닝, 양자화 등 모델압축기술, 학습과 추론 속도를 높이고 시스템 자원 활용 개선을 위한 커널 최적화 등 현 트랜스포머 아키텍처 내에서 개선 연구와 더불어 SSMs(State Space Models)나 MoE(Mixture of Experts)와 같은 차세대 아키텍처에 대한 연구도 병행
  - 산업 표준 확립을 통한 기술 호환성 보장, 다양한 환경에서의 성능 검증을 위한 벤치마크 개발, 개발자 성능 향상을 위한 도구 및 프레임워크 개선, 효율적인 모델 배포와 운영을 위한 시스템 구축 등 실용화 연구도 필요

- 대규모 학습 및 추론 최적화 기술
  - 분산 데이터 병렬, 텐서 병렬화, 파이프라인 병렬화 등 병렬 컴퓨팅 기술, 그래디언트 체크포인트, CPU 오프로딩, KV 캐시(Key-Value Cache) 최적화 등 메모리 관리 기술, LoRA(Low-Rank Adaptation), 프롬프트 튜닝 등 효율적인 파라미터 학습방법 등이 연구 중
  - 향후에는 지연시간과 처리율의 균형을 맞추는 최적화 기술, 복잡한 분산 시스템에서 성능 병목현상을 자동으로 감지하고 해결하는 기술, 리소스 할당을 자동화하는 기술, 하드웨어-소프트웨어 통합 최적화, 학습과 추론에 소요되는 에너지 소비를 줄이면서 성능을 유지할 수 있는 에너지 효율성 향상 등 연구 필요



□ (안전과 신뢰) 안전성과 신뢰성 확보, 인간 가치를 정렬하기 위한 방법론 등

○ 안전성 및 신뢰성 확보

- 적대적 공격 방어 메카니즘, 모델 출력 신뢰도 평가 및 불확실성 추정, 의사결정 검증 시스템 구축 등의 안전성 보장 기술, 의사결정 과정 해석기술 개발, 결과에 대한 근거 제시 등의 설명 가능성과 투명성 연구, 연합학습을 통한 데이터 프라이버시 보호, 안전한 데이터 공유 프레임 구축 등의 프라이버시 보호 연구 등 수행 중
- 향후에는 초인 수준의 AGI 시스템에 대한 안전성 보장, 장기적 영향을 고려한 안전성 평가 방법론, 복잡한 시스템 상호작용에서의 의사결정 방법론 등 연구 필요

○ 인간 가치를 정렬하기 위한 윤리적 추론 및 의사결정 능력 개발

- RLHF(Reinforcement Learning from Human Feedback)를 통한 인간 가치 학습, 헌법적 AI 방법론을 통한 윤리적 원칙 내재화, 다양한 문화와 가치를 고려한 포용적 학습 데이터 구축, 편향성과 차별을 줄이기 위한 공정성 향상 기술 등 연구
- \* 헌법적 AI(Constitutional AI) 방법론이란 헌법의 기본 원칙과 가치, 예를 들어 인간의 존엄성, 프라이버시 보호, 평등권 등의 헌법적 가치를 AI 개발과 사용에 반영하여 윤리적 문제를 해결하는 접근 방식
- 문화적 다양성을 고려한 윤리적 의사결정 체계, 동적으로 변화하는 가치체계 수용 방안, 사회적 합의를 반영한 윤리 가이드라인 개발 등 필요

□ (응용) 과학을 위한 AI, 소프트웨어 자동생성 기술, 다양한 사회 문제해결 분야 등

○ 과학 연구 및 발견

- (생명의학) 단백질 구조 예측, 신약 개발, 질병 진단 등
- (물리학) 양자 시스템 연구, 천체 물리학 데이터 분석, 중력파 검출 등
- (수학) 정리 증명, 수학적 패턴 발견, 복잡한 문제해결 등
- (화학) 분자 구조 예측, 새로운 물질 설계, 반응 경로 분석 등

○ 코딩 및 소프트웨어 개발

- 자동코드 생성, 코드 최적화, 소프트웨어 테스트, AI 기반 개발자 생산성 향상 연구 등

○ 이외, 기후, 우주탐사, 교육, 의료, 헬스케어, 사회 문제해결 등 다양한 분야에서 응용

## 4 빅테크 기업의 AGI 연구 및 개발 방향

### 1) 기업 선정 근거와 연구 영향력 분석

#### □ 주요 기업 선정 근거

○ 2024 AI Index 보고서<sup>22)</sup>의 AGI 관련 성능 평가에서 우수한 성능을 달성한 기관을 우선 선택하고, 그 외 AGI 개발 성과가 뉴스에 보도된 기업들을 추가

- AI Index 보고서의 Expert AGI를 평가하기 위한 MMMU 벤치마크에서 50% 이상의 성능을 달성한 구글, 오픈AI, 상하이AI연구소, 알리바바, 마이크로소프트 고려

\* MMMU(Massive Multi-discipline Multimodal Understanding)는 6개의 도메인(Art and Design, Business, Science, Health and Medicine, Humanities and Social Science, Technology and Engineering)에서 만들어진 11,500개의 대학생 수준의 질문으로 구성

#### [2024년 AI Index의 MMMU 평가<sup>23)</sup> 결과]

모델명	성능	기관명	분류	국가
Gemini Ultra	59.4%	구글	기업	미국
GPT-4V	56.8%	오픈AI	기업	미국
InternVL-MAX	51.6%	상하이AI 연구소	연구소	중국
Qwen-VL-Chat-V1.2	51.4%	알리바바	기업	중국
LLaVA-1.6-34B	51.1%	마이크로소프트외	기업	미국

- AI Index 보고서의 AI 윤리와 정렬(alignment) 평가에서 우수한 성능 달성한 오픈 AI, 앤스로픽, 스탠포드대학 고려

\* AI 모델이 별도의 학습 없이(zero-shot) 도덕적 허용 가능성을 판단하는 과제에서 인간의 판단과 일치하는지를 완전 일치 여부(discrete agreement)로 측정된 결과

#### [2024년 AI Index의 Zero-shot Alignment<sup>24)</sup> 평가 결과]

모델명	성능	기관명	분류	국가
GPT-4	41.9	오픈AI	기업	미국
GPT-3.5-Turbo	40.3	오픈AI	기업	미국
Claude-v1	37.1	앤트로픽	기업	미국
Alpaca-7B	32.3	스탠포드대학	대학	미국

22) 스탠포드대 인간중심인공지능연구소(HAI: Human-Centered Artificial Intelligence), Artificial Intelligence Index Report (2024.4)

23) 전문가 수준의 AGI를 평가하기 위해 최근에 새롭게 구축된 MMMU(Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI) 벤치마크 셋으로 평가

24) 윤리적 의사결정과 도덕적 판단을 평가하기 위해서 스탠포드에서 개발한 데이터셋 MoCA(Moral Choice Alignment)로 평가

- 최근 AGI 개발성으로 AlphaGemetry, V-JEPA, GPT-o1이 보도된 **딥마인드, 메타, 오픈 AI**를 고려
- \* 딥마인드가 공개한(2024.8) AlphaGemetry2는 국제 수학 올림피아드 문제에서 은메달 수준으로 문제를 풀이 함으로써 복잡한 논리적 문제를 해결할 수 있는 AI의 가능성을 보여줌
- \* 메타가 공개한(2024.4) V-JEPA는 트랜스포머 아키텍처를 기반으로 하는 기존의 대형언어모델과 다른 인간의 학습 방식을 모방하는 AI 모델로 자율학습을 통해 세계에 대한 지식을 획득하는 것을 목표로 함
- \* 오픈AI는 AGI 개발의 5단계 중 2단계에 근접한 추론 성능을 보여주는 GPT-o1 출시(24.9)<sup>25)</sup>
- 이 중 AGI 관련 주요 개발자의 인터뷰 및 논문 등의 자료 확보가 가능한 **구글, 딥마인드, 오픈AI, 메타, 앤스로픽**의 연구 방향 분석

□ 5개 기업의 연구 영향력

- 구글 스칼라에서 제시하는 H5-index 기준 인공지능과 컴퓨터비전 및 패턴인식 분야의 상위 5개 학회에서 5대 기업 - **구글, 딥마인드, 오픈AI, 메타, 앤스로픽** - 연구 영향력 분석
- Engineering & Computer Science 분야의 하위 분야인 AI와 CVPR에서 H5-index 기준 상위 5개 학회에서 논문 수집<sup>26)</sup>
  - \* 구글 스칼라에서 학회 및 저널을 평가하기 위해서 사용하는 지표로 최근 5년 해당 학회에 제출된 모든 논문들의 H-index를 기반으로 계산
- H5-index 기준으로 CVPR, NeurIPS, ICLR, ICCV, ICML을 선별
- \* 저널보다는 학회 활동이 활발한 AI분야의 특성을 고려하고 최근 연구동향을 살펴보고자 H5-index 기준 상위 랭크에 위치하더라도 저널은 제외

[AI와 CVPR 분야에서 H5-index 기준 상위 5개 학회 리스트]

Engineering & Computer Science 분야	학회명	H5-index
Computer Vision and Pattern Recognition	Conference on Computer Vision and Pattern Recognition (CVPR)	440
Artificial Intelligence	Neural Information Processing Systems (NeurIPS)	337
Artificial Intelligence	International Conference on Learning Representations (ICLR)	304
Computer Vision and Pattern Recognition	International Conference on Computer Vision (ICCV)	291
Artificial Intelligence	International Conference on Machine Learning (ICML)	268

- (5개 빅테크 기업 전체) 피인용 수 기준 상위 1% 논문에서 차지하는 빅테크 기업들의 논문 수 비중은 연 기준 최소 15% 이상이며 피인용 수 총합은 이보다 높은 최소 17% 이상

25) Introducing OpenAI o1-preview (2024.9.12.) (출처: <https://openai.com/index/introducing-openai-o1-preview/>)

26) 논문데이터베이스 서비스 웹 사이트인 Scopus에서 2024년 10월 수집

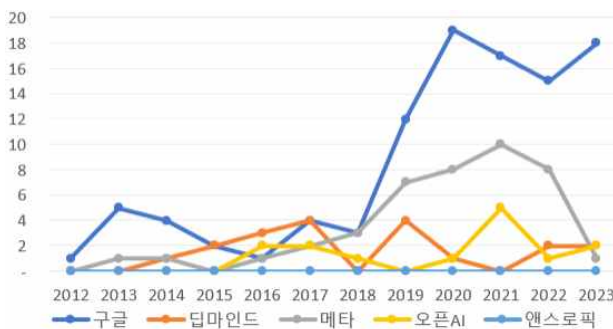
- 5개 기업의 전체 논문 수 비중은 2017년까지 대체로 50% 내외로 매우 높았으나, 최근 5년 하락하여 2023년도 15% 수준으로 감소
- 피인용 수는 상위 1% 논문에서 이들 기업들이 차지하는 비중은 더욱 높아 연구에서 이들의 영향력이 매우 크다는 것을 보여줌



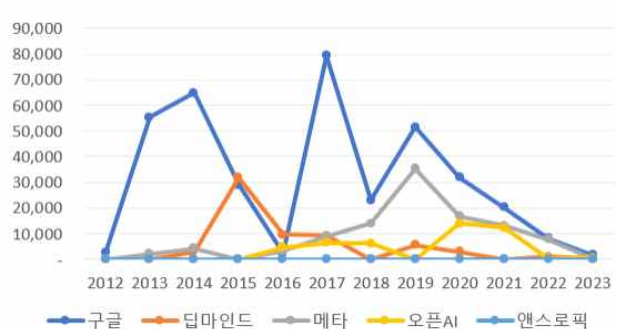
[피인용 수 기준 상위 1% 논문에서 5개 빅테크 기업의 논문 수 및 피인용 수 총합 비중]

o (기업별) 피인용 수 기준 상위 1% 논문 수 및 피인용 수 총합 비교

- (구글) 2017년 출판된 트랜스포머가 소개된 논문의 지대한 영향으로 2017년 피인용수에서 정점을 찍은 구글은 이후 논문 출판 건수도 크게 증가
  - \* “Attention is all you need”에서 현 대형언어모델의 주류 아키텍처인 트랜스포머 소개
- (답마인드) 2015년 답마인드 논문의 피인용 수 총합이 매우 증가했으며, 이는 알파고 소개 이후 연구 영향력이 확대되었음을 의미
- (메타, 오픈AI) 메타는 2019년, 오픈AI는 2020년 피인용 수 총합이 크게 증가했으며, 오픈AI의 경우 이 시기에 발표된 GPT-3 논문의 영향
  - \* 메타의 경우 2019년 피인용 수 증가는 구글과 함께 출판한 오픈소스 머신러닝 라이브러리 Pytorch를 소개한 논문 영향
- (앤스로픽) 2021년에 설립된 회사로 논문성과를 분석하기에 아직은 시기상조



[기업별 상위 1% 논문 수]



[기업별 상위 1% 논문의 피인용 수 총합]

- 2012년부터 2023년까지의 연도별 피인용 수 상위 5개의 논문을 살펴본 결과, 2010년대는 구글이 주도하다가, 2020년부터 오픈AI의 영향력 증대
- 2010년대 구글이 연구 생태계를 주도하다가 2020년부터 오픈AI의 영향력 증대되어, 현재 오픈AI, 구글, 메타가 주도 중

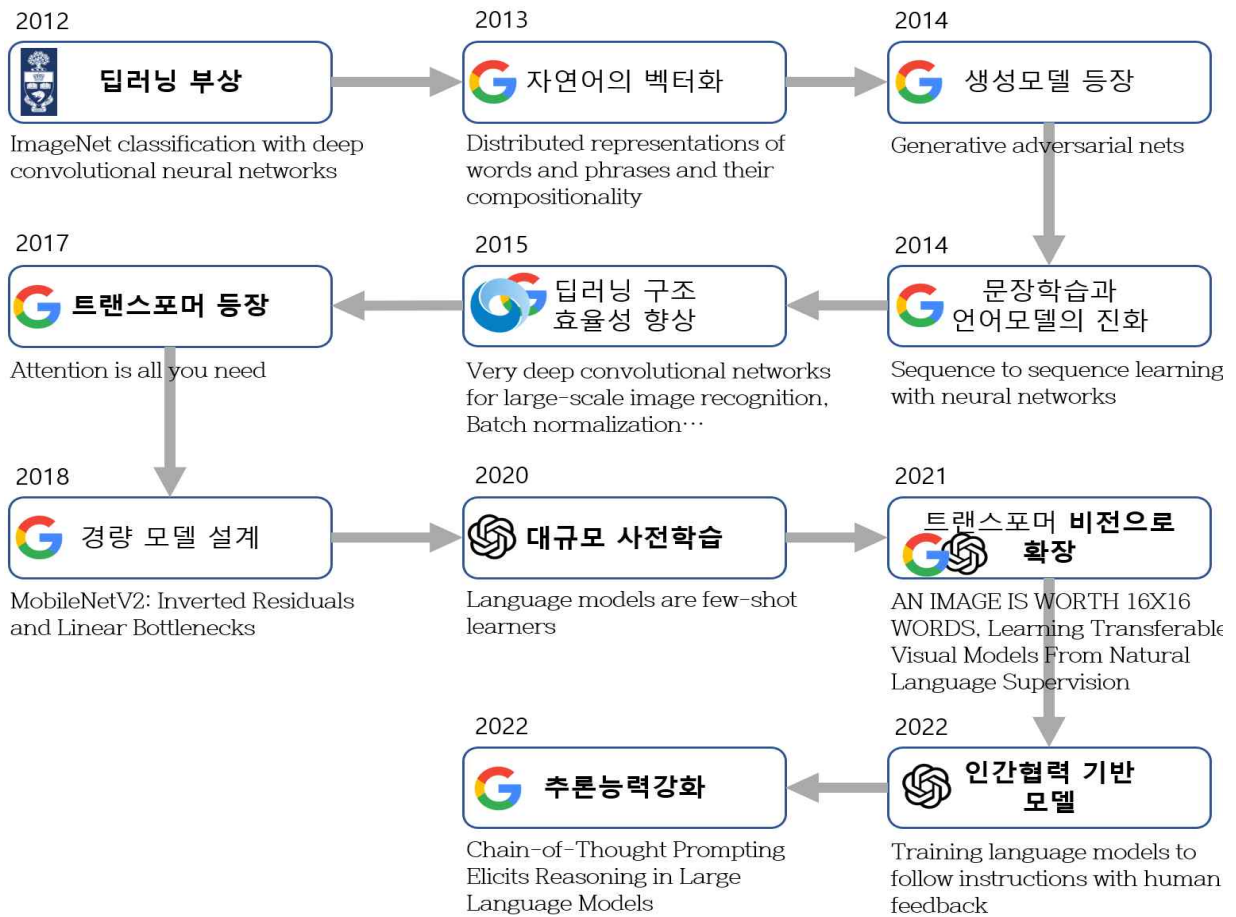
[피인용수 기준 연도별 상위 5개 논문 리스트]

연도	논문 제목	피인용수	대표기관
2012	<b>ImageNet classification with deep convolutional neural networks</b>	81,221	토론토대
	Practical Bayesian optimization of machine learning algorithms	5,157	토론토대
	Large scale distributed deep networks	2,590	구글
	Learning to discover social circles in ego networks	1,358	-
	Image denoising and inpainting with deep neural networks	1,222	USTC
2013	<b>Distributed representations of words and phrases and their compositionality</b>	24,250	구글
	<b>Efficient estimation of word representations in vector space</b>	21,557	구글
	Translating embeddings for modeling multi-relational data	6,350	CNRS
	Sinkhorn distances: Lightspeed computation of optimal transport	2,395	교토대
	DeViSE: A deep visual-semantic embedding model	2,017	구글
2014	<b>Generative adversarial nets</b>	44,109	구글
	<b>Sequence to sequence learning with neural networks</b>	14,406	구글
	Auto-encoding variational bayes	8,997	암스테르담대
	Two-stream convolutional networks for action recognition in videos	5,929	코넬대
	How transferable are features in deep neural networks?	5,835	옥스포드대
2015	Adam: A method for stochastic optimization	48,153	MS
	Faster R-CNN: Towards real-time object detection with region proposal networks	30,513	암스테르담대
	<b>Very deep convolutional networks for large-scale image recognition</b>	26,195	답마인드
	<b>Batch normalization: Accelerating deep network training by reducing internal covariate shift</b>	23,774	구글
	Neural machine translation by jointly learning to align and translate	8,927	Jacobs대
2016	Convolutional neural networks on graphs with fast localized spectral filtering	6,000	EPFL
	Improved techniques for training GANs	5,514	-
	Matching networks for one shot learning	5,065	답마인드
	R-FCN: Object detection via region-based fully convolutional networks	4,823	메타
	Unsupervised representation learning with deep convolutional generative adversarial networks	3,240	답마인드
2017	<b>Attention is all you need</b>	67,034	구글
	Semi-supervised classification with graph convolutional networks	12,753	암스테르담대
	A unified approach to interpreting model predictions	12,753	워싱턴대

	Inductive representation learning on large graphs	9,828	스탠포드대
	LightGBM: A highly efficient gradient boosting decision tree	7,384	MS
2018	Squeeze-and-Excitation Networks	16,802	Momenta
	<b>MobileNetV2: Inverted Residuals and Linear Bottlenecks</b>	<b>8,682</b>	구글
	Non-local Neural Networks	6,255	CMU
	The Unreasonable Effectiveness of Deep Features as a Perceptual Metric	6,155	UC 버클리
	Path Aggregation Network for Instance Segmentation	5,934	홍콩대
2019	PyTorch: An imperative style, high-performance deep learning library	24,877	구글/메타
	Searching for mobileNetV3	5,864	구글
	A style-based generator architecture for generative adversarial networks	5,791	NVIDIA
	Dual attention network for scene segmentation	4,861	CAS
	XLNet: Generalized autoregressive pretraining for language understanding	4,658	CMU
2020	<b>Language models are few-shot learners</b>	<b>13,939</b>	<b>오픈AI</b>
	Momentum Contrast for Unsupervised Visual Representation Learning	7656	메타
	A simple framework for contrastive learning of visual representations	6,075	구글
	Denosing diffusion probabilistic models	4,875	UC 버클리
	EfficientDet: Scalable and efficient object detection	4,774	구글
2021	Swin Transformer: Hierarchical Vision Transformer using Shifted Windows	12,070	MS
	<b>AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE</b>	<b>7,521</b>	구글
	<b>Learning Transferable Visual Models From Natural Language Supervision</b>	<b>6926</b>	<b>오픈AI</b>
	Training data-efficient image transformers & distillation through attention	3193	메타
	Coordinate attention for efficient mobile network design	3148	싱가포르대
2022	High-Resolution Image Synthesis with Latent Diffusion Models	3737	Fudan대
	<b>Training language models to follow instructions with human feedback</b>	<b>2937</b>	<b>오픈AI</b>
	A ConvNet for the 2020s	2810	메타
	Masked Autoencoders Are Scalable Vision Learners	2680	메타
	<b>Chain-of-Thought Prompting Elicits Reasoning in Large Language Models</b>	<b>2215</b>	구글
2023	Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks	560	Hkust
	Robust Speech Recognition via Large-Scale Weak Supervision	457	오픈AI
	Segment Anything	427	메타
	Adding Conditional Control to Text-to-Image Diffusion Models	349	스탠포드대
	BiFormer: Vision Transformer with Bi-Level Routing Attention	337	홍콩대



- 피인용 수 상위 5개 논문 중 주요 빅테크 기업의 AI 논문성과는 AI 기술 발전을 보여줌
  - (2012년~2013년, 딥러닝 부상) ImageNet 대회에서 압도적인 성능을 보여주며 2012년 등장한 AlexNet은 AI모델들이 딥러닝으로 전환되는 계기를 마련했으며, 2013년 Mikolov와 연구진이 발표한 Word2Vec은 지금의 자연어처리의 기초가 됨
  - (2014년~2017년, 아키텍처 발전) 2014년 Goodfellow와 동료들이 발표한 GAN은 AI가 실제와 유사한 데이터를 생성할 수 있다는 가능성을 보여주었고, 그 해 구글이 발표한 Sequence to Sequence 학습방법은 기계번역의 시대를 열었으며, 2017년 트랜스포머 아키텍처 논문이 발표되면서 현대 AI의 가장 큰 전환점을 가져옴
  - (2018년~2020년, 대규모언어모델 시대) 2018년 MobileNetV2가 등장하여 모바일 기기에서도 효율적인 AI 구현이 가능해졌으며, 2020년 OpenAI가 발표한 GPT-3으로 대규모 언어모델의 시대가 개화



[최근 10년 피인용기준 Top 논문들로부터 살펴본 AI 발전]

- (2021년~, 멀티모달과 추론능력 강화) 2021년 구글이 발표한 ViT(Vision Transformer)와 OpenAI의 CLIP으로 컴퓨터비전과 자연어처리의 통합이 가속화되었고, 2022년 InstructGPT를 통해 AI의 안전성과 유용성을 높이는 RLHF가 도입되었으며, 구글의 Chain of Thought 프롬프팅은 AI 추론능력을 크게 향상시킴

## 2) 5개 기업의 AGI 연구 및 개발 방향

### □ 구글

- 現 대형언어모델의 주류 아키텍처인 트랜스포머(Transformer)를 개발한 구글은 언어 모델 고도화, 멀티모달학습 강화, 추론능력 개선이라는 소프트웨어 혁신과 함께 AI 전용 칩과 양자컴퓨팅이라는 하드웨어 혁신을 동시에 추구하면서 보다 효율적이고 범용적인 AGI 개발을 위한 통합적 접근을 시도 중
- Word2Vec 모델의 기초 위에, Seq to Seq Learning을 거쳐 Transformer 아키텍처를 개발했으며, Switch Transformer를 통한 효율적 학습, LaMDA의 대화 특화, PaLM의 540B 매개변수 규모 확장, BERT의 양방향 문맥이해 등 대규모언어모델에서 지속적인 혁신을 시도
- 멀티모달학습에서는 ViT를 통해 이미지에 트랜스포머 적용, Imagen을 통한 이미지와 텍스트 통합 이해, AudioLM을 통한 음성 및 음향 효과 생성, Gemini 시리즈의 통합적 처리, Pathways를 통한 범용성과 효율적 아키텍처 제시
- 추론능력 향상을 위해서는 Chain-of-thought Prompting과 신경망과 심볼릭 AI를 결합한 하이브리드 접근 등 다양한 연구 시도 중
- 하드웨어 측면에서는 TPU6 개발과 AI 기반 칩 설계 최적화를 진행하는 동시에 양자 컴퓨터를 활용한 새로운 기계학습 알고리즘 개발과 양자정보처리프로세서(QPU) 시스템 연구도 수행하고 있음
- 이러한 소프트웨어와 하드웨어의 양방향 접근을 통해 AGI 개발을 가속화하고 있으며, AutoML과 MobileNetVx을 통한 시스템 최적화 연구로 실용성도 함께 고려 중
- 또한 로봇공학 분야의 RT-X 프로젝트, 로봇에 제미나이 결합, 실시간 산불 식별 및 추적 시스템, 홍수예보와 낙후된 지역 원격의료시스템 등 다양한 실세계 문제해결에 AI를 적용 연구도 수행 중

### [AGI 연구 성과 및 방향 - 구글]

	주요 연구	연구 성과 및 방향
1	대규모언어모델	<ul style="list-style-type: none"> <li>- 단어와 구문을 벡터공간에 임베딩하는 방법 제안 Word2Vec 모델의 기초</li> <li>- 생성모델과 판별모델이 경쟁하는 적대적 학습구조 GAN</li> <li>- 인코더와 디코더로 구성된 자연어처리 대표모델 Seq to Seq Learning</li> <li>- 현 대규모언어모델의 주류 아키텍처 Transformer</li> <li>- 대규모언어모델 효율적인 학습을 위한 Switch Transformer 모델</li> <li>- 대화에 특화된 LaMDA</li> <li>- 540B 매개변수를 갖고 있는 대규모 언어모델 PaLM</li> <li>- 양방향 문맥이해 모델 BERT</li> </ul>
2	멀티모달 학습	<ul style="list-style-type: none"> <li>- 이미지 인식에 트랜스포머 아키텍처 활용 ViT(Vision Transformer)</li> <li>- 이미지와 텍스트를 동시에 이해하는 모델 Imagen</li> </ul>



		<ul style="list-style-type: none"> <li>- 동영상에 맞는 음성 및 음향 효과를 생성하는 오디오 모델 <b>AudioLM</b></li> <li>- 멀티모달 대규모모델 <b>Gemini</b> 시리즈</li> <li>- 범용성, 일반화, 효율적 학습이 가능한 AI 모델 아키텍처 <b>Pathways</b></li> <li>- 테스트를 이미지로 생성하는 모델 <b>Veo</b></li> </ul>
3	계획 및 추론	<ul style="list-style-type: none"> <li>- 규칙을 모르는 상태에서 게임을 학습하는 강화 알고리즘 연구</li> <li>- 복잡한 추론과정을 단계별로 생성하는 기술 <b>Chain-of-Thought Prompting</b></li> </ul>
4	메모리	<ul style="list-style-type: none"> <li>- 시간패턴을 학습하고 예측에 유용한 인간의 신피질을 모방한 모델 <b>HTM</b></li> <li>- 트랜스포머 아키텍처의 어텐션 메커니즘을 활용하여 단기기억을 구현</li> <li>- 대규모 지식 그래프와 데이터베이스를 신경망과 결합</li> </ul>
5	메타 학습	<ul style="list-style-type: none"> <li>- 다양한 방법론에 적용 가능한 메타학습 제시, <b>MAML</b></li> <li>- 로봇이 인간처럼 학습, 적응, 진화하는 방법 특허 출원, <b>Meta learning System</b></li> </ul>
6	에이전트	<ul style="list-style-type: none"> <li>- 카메라를 통해 파악한 사물 설명과 주변위치 등 파악, <b>아스트라</b></li> <li>- 웹브라우저를 작동해 검색, 제품구매, 항공편예약 등 수행, <b>자비스<sup>27)</sup></b></li> </ul>
7	시스템 효율화	<ul style="list-style-type: none"> <li>- 모바일 환경에서의 딥러닝 모델 활용성 확장 <b>MobileNetVx</b></li> </ul>
8	자동화된 기계학습	<ul style="list-style-type: none"> <li>- 최적의 신경망 구조를 자동으로 탐색 모델 <b>AutoML</b></li> <li>- 모델의 자동설계 및 최적화에 활용</li> </ul>
9	양자컴퓨터와 AI	<ul style="list-style-type: none"> <li>- 양자컴퓨터를 활용한 새로운 기계학습 알고리즘 개발</li> <li>- 양자컴퓨터와 고전컴퓨터를 결합한 AI 시스템 연구</li> <li>- 양자정보처리프로세서(QPU) 시커모어</li> <li>- 오류정정이 가능한 양자칩 <b>윌로우(QPU)<sup>28)</sup></b></li> </ul>
10	AI칩 설계 및 개발	<ul style="list-style-type: none"> <li>- AI전용 하드웨어 <b>TPU6</b></li> <li>- AI를 사용하여 칩설계 최적화</li> </ul>
11	로봇공학	<ul style="list-style-type: none"> <li>- 범용 로봇 브레인 개발 <b>RT-X 프로젝트</b></li> <li>- 주변환경을 학습해 명령을 수행할 수 있도록 AI로봇에 제미니가 결합</li> <li>- 사람처럼 상호작용하는 길 안내 로봇</li> </ul>
12	실세계 문제해결	<ul style="list-style-type: none"> <li>- 실시간을 산불을 식별하고 추적하는 AI 시스템</li> <li>- 홍수예보를 위해 AI 활용</li> <li>- 낙후된 지역을 위한 원격의료시스템에 AI 활용</li> </ul>
13	AI 안전성 및 윤리	<ul style="list-style-type: none"> <li>- 구글의 AI 원칙 수립 및 적용</li> <li>- 모델의 편향성을 줄이기 위한 연구</li> </ul>

\* 주요 논문성과, 뉴스보도, 주요인사의 인터뷰 기반으로 정리

- (주요 연구 분야<sup>29)</sup>) 머신러닝, 딥러닝, 컴퓨터비전, 인간-컴퓨터 상호작용(HCI) 등 인공지능 및 컴퓨터과학 분야에서 연구성과가 주로 도출되고 있으며 생물정보학과 같은 특화된 분야에서 AI 기술을 적용한 연구가 활발히 진행 중

27) 구글, AI 에이전트 ‘자비스’ 12월 공개 예정 (2024.10.27.) (출처: <https://www.aitimes.com/news/articleView.html?idxno=164658>)

28) <https://blog.google/intl/ko-kr/company-news/technology/quantum-ai-willow-kr/>

29) 딥러닝으로 인공지능 연구의 패러다임 전환을 이끌었던 AlexNet이 출시된 2012년 이후, 논문 게재를 가장 많이 하고 있는 상위 5개의 학회리스트로 부터 유추

[주요 연구 분야 - 구글]

학회 및 저널명	주요 세부 영역	논문수
Advances in Neural Information Processing Systems	딥러닝 및 머신러닝, 강화학습, 신경망아키텍처, 인공지능 이론	1,008
Proceedings of Machine Learning Research	기계학습, 통계적학습이론, 비지도 학습 및 지도학습, 인공지능 및 로보틱스	832
Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	인공지능, 지식표현 및 추론, 자연어처리, 다중 에이전트 시스템, 생물정보학, 유전체학 및 단백질 구조분석, 의료AI, 시뮬레이션 및 모델링	721
Proceedings of the IEEE computer Society Conference on Computer Vision and Pattern Recognition	컴퓨터비전, 패턴인식, 3D 컴퓨터 비전, 영상 생성 및 변환	429
Conference on Human Factors in Computing Systems Proceedings	인간-컴퓨터 상호작용(HCI), 사용자연구, 유저경험(UX)디자인, 사회적 영향 연구	321

\* 2013년 이후, 총 17,075건

- (주요 협력관계<sup>30)</sup>) 상위 10개의 기관 중 자국내 위치한 9개의 기관과 주로 협력하고 있으며, 대부분 대학과 협력 중

[주요 협력관계 - 구글]

협력 기관명	협력 논문수	비고
Stanford University	833	미국
University of California, Berkeley	810	미국
Carnegie Mellon University	638	미국
Massachusetts Institute of Technology	586	미국
Columbia University	389	미국
Harvard University	350	미국
Princeton University	331	미국
Georgia Institute of Technology	330	미국
Microsoft Research	329	미국, 기업
Tel Aviv University	313	이스라엘

- (주요 국내 협력관계) 공동저자 논문 수 기준 상위 협력기관은 연세대, KAIST, 고려대 등 대학들이며, 연구원 중에서는 한국천문연구원과 협력 중

\* 천문연구원과는 과학분야에서의 AI 활용의 한 예로 우주(Galaxies), 블랙홀 등의 연구 진행

30) 공동 논문저자의 소속기관으로 협력관계 파악

[국내 기관과의 협력관계 - 구글]

협력 기관명	협력 논문수	비고
연세대학교	94	대학
서울대학교	85	대학
KAIST	72	대학
고려대	41	대학
한국천문연구원	26	연구원

\* 2013년 이후, 총 353건

□ 딥마인드

- 뛰어난 강화학습 능력을 기반으로 알파고, 알파제로, 알파폴드 등 특정 도메인에서 인간 수준을 능가하는 AGI를 개발한 딥마인드는 최근에 AGI 5단계 초인(Super Intelligence) 수준의 지능을 달성하기 위해서는 오픈엔드 학습 중요성을 강조하며 오픈엔드 기초 모델 개발 방향<sup>31)</sup>을 제시하고 있으며, 특히 수학적 추론과 과학적 문제해결 문제에서 주목할 만한 성과를 도출

\* 오픈엔드 학습(open\_ended learning)은 관찰자의 관점에서 새로움(novelty)과 학습가능성(learnability)을 지속적으로 생성하는 학습개념이며, AGI가 필수적으로 갖추어야 할 자율성장과 연관

- 280B 매개변수를 가진 Gopher와 7B 규모의 Chinchilla 등 대규모언어모델을 개발하며 AGI의 기본모델의 기반을 다지고, AlphaZero, AlphaProof, AlphaGeometry2와 같이 규칙 기반의 게임과 수학 분야에서 뛰어난 성과를 보이며, ‘Reward is Enough’ 원칙하에 강화학습을 통한 AGI 발전을 추구
- 멀티모달 측면에서는 Perceiver와 Flamingo를 통해 다양한 입력 모달리티를 처리하는 범용 아키텍처를 개발했으며, XLand 환경에서의 3D 시뮬레이션을 통해 복잡하고 다양한 과제 해결 능력을 향상
- 또한 Neuro-symbolic AI 연구를 통해 신경과학과 AI의 융합 연구도 시도
- 최근에는 오픈엔드 학습을 통해 XLand 환경에서 훈련된 유능한 에이전트를 개발하고, 단백질 구조 예측 모델 AlphaFold와 과학적 문제해결을 위한 AI for Science 등을 통해 실세계 문제해결에서 성과를 도출
- SynthID를 통한 AI 생성 이미지 워터마크 연구와 AI 시스템의 예측 불가능한 행동변화 연구 등 AI 안전성과 윤리적 측면도 함께 연구 중

31) Open-Endedness is Essential for Artificial Superhuman Intelligence, arXiv:2406.04268 (2024.6.6.)

[AGI 연구 성과 및 방향 - 딥마인드]

	주요 연구	연구 성과 및 방향
1	대규모언어모델	- 280B 매개변수를 갖고 있는 대규모 언어모델 Gopher - 7B 모델 Chinchilla로 계산최적모델 제시 - 단일 모델로 다양한 작업을 수행하는 멀티태스크 에이전트 Gato
2	멀티모달 학습	- 다양한 입력 모달리티를 처리할 수 있는 범용 아키텍처 Perceiver - 텍스트와 이미지를 통합한 멀티모달모델 Flamingo
3	계획 및 추론	- 규칙만으로 다양한 게임을 마스터하는 일반 강화학습 모델 AlphaZero - 보상 최대화 원칙만으로 구성된 학습 프레임워크 제안 Reward is Enough - 복잡하고 다양한 과제를 제공하는 3D 시뮬레이션 XLand - 수학추론에 특화된 강화학습 AI 모델 AlphaProof - 국제수학올림피아드에서 은메달 수준의 성능을 달성한 모델 AlphaGeometry2 - 고품질 합성데이터로 미세조정된 모델로 추론성능 개선 - MoE를 수백만 전문가로 확장하는 새로운 아키텍처 PEER(Parameter Efficient Expert Retrieval)
4	메모리	- 외부메모리를 사용하는 신경망 모델 RETRO - 지속적인 학습 모델 Elastic Weight Consolidation
5	메타 학습	- 유니버설 튜링 머신(UTM)을 사용하여 다양한 패턴의 훈련데이터 생성
6	에이전트	- 시스템2 사고 강화한 AI 에이전트 프레임워크 Talker-Reasoner
7	신경과학과 AI	- 신경과학과 AI연구의 상호이익적 관계 탐구 Neuro-symbolic AI - 실제 쥐에서 기록한 고해상도 데이터로 인공신경망 훈련
8	오픈엔드 시스템	- 오픈엔드 학습으로 XLand 환경에서 훈련된 유능한 에이전트 개발 - 오픈엔드 기초모델 개발방향 제시
9	로봇공학	- 대형언어모델으로 로봇지능 고도화 오토RT - 아마추어 인간 수준의 탁구로봇 개발 - 신발 끈 묶는 로봇 핸드 기술 개발
10	실세계 문제해결	- 단백질 구조 예측 모델 AlphaFold - RL 알고리즘 기반 행렬곱 연산속도 개선 모델 AlphaTensor - 기후변화, 에너지효율성 등 과학적 문제해결을 위한 AI for Science
11	AI 안전성 및 윤리	- AI 생성 이미지에 픽셀 단위로 투명한 워터마크를 삽입 도구, SynthID - 학습된 AI 시스템의 예기치 않는 행동변화 연구

\* 주요 논문성과, 뉴스기사, 주요인사의 인터뷰 기반으로 정리

- (주요 연구 분야) 구글 스칼라에서 제시하는 H5-index 기준 인공지능과 컴퓨터비전 및 패턴인식 분야의 상위에 위치한 우수학회 및 저널에서의 연구활동이 특히 활발하며, 과학분야에서의 AI 활용에서 상당한 연구성과 도출

[주요 연구 분야 - 딥마인드]

학회 및 저널명	주요 세부 영역	논문수
Proceedings Of Machine Learning Research	기계학습, 통계적학습이론, 비지도 학습 및 지도학습, 인공지능 및 로봇틱스	425
Advances In Neural Information Processing Systems	딥러닝 및 머신러닝, 강화학습, 신경망아키텍처, 인공지능 이론	356
International Conference On Learning Representations	자기지도학습, 생성 모델, 강화학습, 신경망 구조	223
International Conference On Machine Learning	지도학습, 비지도학습, 준지도학습, 온라인 학습, 베이저안 방법론	165
Nature	AI/ML 관련 세부 주제: 인공지능의 과학적 응용 (예: 생물학, 화학, 물리학에서의 AI 활용), 뇌과학과 AI의 접점, 양자 컴퓨팅과 AI, AI의 윤리적·사회적 영향	108

\* 2013년 이후, 총 2,380건

- (주요 협력관계) 상위 10곳의 협력기관이 모두 대학이며, 이 중 4곳은 자국 내 위치하며, 그 외 캐나다 대학 1곳과 미국 대학 5곳 등 AI를 선도하는 미국과 활발하게 연구협력 중

[주요 협력관계 - 딥마인드]

협력 기관명	협력 논문수	비고
University of Oxford	228	영국
University College London	217	영국
University of Alberta	128	캐나다
University of California, Berkeley	105	미국
University of Cambridge	103	영국
Princeton University	101	미국
Stanford University	101	미국
Massachusetts Institute of Technology	94	미국
Carnegie Mellon University	88	미국
Imperial College London	74	영국

- (주요 국내 협력관계) KAIST, 고려대, 서울대, 포항공대 등 대학들과 협력하고 있으며, 연구원 중에서는 ETRI와 협력

[국내 기관과의 협력관계 - 딥마인드]

협력 기관명	협력 논문수	비고
KAIST	14	대학
고려대학교	8	대학
서울대학교	7	대학
ETRI	4	연구기관
포항공대	3	대학

\* 2013년 이후, 총 34건

□ 오픈AI

- GPT 시리즈에서 시작하여 **최근 출시된 o1 모델까지 언어모델의 성능을 획기적으로 발전시켰으며** 멀티모달능력 강화, 플러그인을 통한 도구 활용성 확장, 콘텐츠 조정 자동화와 성능평가 시스템 구축 등 포괄적이고 실용적인 접근으로 AGI 개발을 추구하고, 자체 정의한 **AGI 2단계의 추론자 수준을 달성하고 에이전트 단계로 진입**

\* 차세대 프론티어 모델 오리온(Orion)은 오픈AI가 정의한 AGI 2단계 추론자 수준<sup>32)</sup>

- ChatGPT 서비스를 출시하며 대규모언어모델 개발을 선도하고 있으며, GPT-2의 대규모 비지도 학습, GPT-3의 few-shot 학습능력 향상, InstructGPT를 통한 인간 피드백 기반의 모델 미세조정 등 지속적인 연구 성과 공개
- 멀티모달학습에서는 CLIP을 통한 텍스트-이미지 관계 학습, DALLE를 통한 이미지 생성, Sora를 통한 비디오생성 등 성과를 도출
- 계획 및 추론 분야에서는 인간의 선호도를 기반으로 하는 강화학습 방법을 제안하고, o1 모델에서 추론능력을 대폭 강화
- 메모리 관리 측면에서는 의미기억과 일화기억의 분리 연구를 진행하고 ChatGPT의 채팅 내용을 장기기억화하는 기술을 개발
- ChatGPT Plugins 시스템을 통해 외부 도구와의 연결성을 확장하고 있으며, 최근에는 Swarm을 통해 에이전트들의 효과적인 협력을 연구하고, 브로드컴 및 TSMC와 협력하여 AI칩 개발도 추진 중
- 한편 윤리 콘텐츠 제거를 위한 콘텐츠조정자동화 시스템을 개발하고, OpenAI Evals를 통해 AI 모델의 성능을 평가하는 등 안전하고 윤리적인 AGI 개발을 위해 노력 중

32) ChatGPT 5: Everything we know so far about Orion, OpenAI's next big LLM (2024.10.29.) (출처 : <https://indianexpress.com/article/technology/artificial-intelligence/openai-chatgpt-5-orion-llm-features-capabilities-launch-9641365/>)

[AGI 연구 성과 및 방향 - 오픈AI]

	주요 연구	연구 성과 및 방향
1	대규모언어모델	- 대형언어모델 GPT 시리즈 개발 - 대규모 비지도 학습을 통한 다중작업 수행 능력 입증, GPT2 - few-shot 학습능력 입증, GPT3 - 인간의 피드백을 통해 언어모델을 미세조정하는 모델 InstructGPT
2	멀티모달 학습	- 텍스트와 이미지 관계 학습방법 CLIP - 텍스트로부터 이미지생성 및 오디오생성 모델 DALE, Sora
3	계획 및 추론	- 인간의 선호도를 기반으로 하는 강화학습 방법 제안 - 추론 기능이 강화된 AI모델 GPT-o1
4	메모리	- 사실기억(semantic memory)과 일화기억(episodic memory)의 분리 연구 - ChatGPT의 채팅내용을 장기기억할 수 있는 메모리 관리 기능 출시
5	메타 학습	- Shortest Descent 알고리즘에 기반한 메타학습 알고리즘, Reptile
6	에이전트	- 에이전트들의 효과적 협력을 위한 멀티 에이전트 프레임워크 Swarm - 사람 대신 컴퓨터를 사용하며 작업할 수 있는 AI 에이전트 개발 <sup>33)</sup> - 자율 머신러닝 엔지니어링 작업을 평가 벤치마크 MLE-bench
7	AI 칩 설계 및 개발	- 미국 브로드컴과 대만 TSMC와 협력하여 AI칩 개발 추진 <sup>34)</sup>
8	로봇공학	- 피규어AI의 로봇 피규어01에 대형멀티모달모델 탑재
9	실세계 문제해결	- 대화형 인공지능 챗봇 ChatGPT - ChatGPT에 외부도구를 연결하는 플러그인 시스템 ChatGPT Plugins - 코드생성 챗봇 Copilot
10	AI 안전성 및 윤리	- 모든 인류에게 이로운 AI 개발을 오픈AI의 미션으로 정의 - GPT-4 기반 유해 콘텐츠 제거를 위한 콘텐츠조정자동화 시스템 - AI 모델의 성능을 평가하기 위한 오픈소스프레임워크 OpenAI Evals

\* 주요 논문성과, 뉴스기사, 주요인사의 인터뷰 기반으로 정리

- (주요 연구 분야) 구글 스칼라에서 제시하는 H5-index 기준 인공지능과 컴퓨터비전 및 패턴인식 분야의 상위 1%에 위치한 우수학회 및 저널에서의 연구활동이 특히 활발하며, 로봇 분야에서도 연구성과 도출 중

33) OpenAI confirms AI agents are coming next year (2024.10.2.) (출처: <https://www.tomsguide.com/ai/chatgpt/the-agents-are-coming-openai-confirms-ai-will-work-without-humans-in-2025>)

34) OpenAI builds first chip with Broadcom and TSMC, scales back foundry ambition (2024.10.31.) (출처: <https://www.reuters.com/technology/artificial-intelligence/openai-builds-first-chip-with-broadcom-tsmc-scales-back-foundry-ambition-2024-10-29/>)

[주요 연구 분야 - 오픈AI]

학회 및 저널명	주요 세부 영역	논문수
Advances In Neural Information Processing Systems	딥러닝 및 머신러닝, 강화학습, 신경망아키텍처, 인공지능 이론	23
International Conference On Learning Representations	자기지도학습, 생성 모델, 강화학습, 신경망 구조	25
Proceedings Of Machine Learning Research	기계학습, 통계적학습이론, 비지도 학습 및 지도학습, 인공지능 및 로봇틱스	17
International Conference On Machine Learning	지도학습, 비지도학습, 준지도학습, 온라인 학습, 베이지안 방법론	14
Proceedings IEEE International Conference On Robotics And Automation	로봇 제어 및 학습, 컴퓨터 비전을 이용한 로봇 인식, 자율 주행, 휴먼-로봇 상호작용, 소프트 로봇틱스, 의료 로봇공학	11

\* 2013년 이후, 총 193건

- (주요 협력관계) 대부분 자국의 대학 및 기업들과 협력 중이며, 영국 및 캐나다의 AI 선도 기관들과도 협력

[주요 협력관계 - 오픈AI]

협력 기관명	협력 논문수	비고
University of California, Berkeley	59	미국
Google LLC	26	미국, 기업
Stanford University	12	미국
DeepMind Technologies Limited	12	영국, 기업
Massachusetts Institute of Technology	9	미국
Carnegie Mellon University	9	미국
University of Toronto	8	캐나다
International Computer Science Institute	8	미국, 연구소
Université McGill	7	캐나다

- (주요 국내 협력관계) UNIST, 성신대학교와 협력

[국내 기관과의 협력관계 - 오픈AI]

협력 기관명	협력 논문수	비고
UNIST	2	대학
성신대학교	1	대학

\* 2013년 이후, 총 353건



□ **메타**

- 효율적인 언어모델 개발, 다중 모달리티의 통합, 메모리 시스템 혁신, 그리고 실세계 응용을 위한 기술 개발을 **오픈소스 전략**을 통해 협력적 발전을 추구하고, **대규모 인프라 투자**를 통해 AGI 개발에 필요한 컴퓨팅 능력도 확보
- \* 2024년 말까지 약 35만 개의 엔비디아 H100 GPU 구매 예정이며, 다른 유사한 GPU를 포함하면 약 60만 개의 칩 보유<sup>35)</sup>
- 175B 규모의 OPT와 소형모델로 대형모델과 유사한 성능을 보이는 LLaMA 시리즈를 오픈소스로 공개하며 효율적인 언어모델 개발을 선도
- 멀티모달 분야에서는 ImageBind를 통해 이미지, 텍스트, 오디오, 깊이, 열, IMU 등 6가지 모달리티의 데이터를 단일 임베딩 공간에 통합하고, V-JEPA를 통해 비디오 자기지도학습을 실현하는 등 다양한 감각 정보의 통합을 추구

[AGI 연구 성과 및 방향 - 메타]

	주요 연구	연구 성과 및 방향
1	대규모언어모델	- 175B 대형언어모델 OPT(오픈소스) - 소형모델로 대형모델과 유사성능을 보이는 <b>LLaMA 시리즈</b> (오픈소스)
2	멀티모달 학습	- 6가지 모달리티(이미지, 텍스트, 오디오, 깊이, 열, IMU(관성측정장비)) 데이터를 단일 임베딩 공간에 통합하는 ImageBind - 비디오를 통한 자기 지도학습 모델 <b>V-JEPA</b> - 테스트로부터 비디오 생성모델 <b>MovieGen</b>
3	계획 및 추론	- 불안정한 정보 게임(포커 등)을 위한 딥강화학습 ReBeL
4	메모리	- 외부메모리를 사용하는 신경망 모델 end-to-end Memory Networks - 지속적인 학습 모델, Gradient Episodic Memory - 메타 챗봇에 메모리 기능 추가
5	로봇공학	- 비전언어모델과 이동계획, 객체조작 등 로봇 기본 요소기술 통합, OK-Robot
6	시스템 효율화	- 텐서수명과 메모리 위치 최적화 알고리즘, MODeL
7	실세계 문제해결	- AI를 핵심요소 기술로 거론하며 <b>메타버스 비전</b> 제시 - <b>AR글래스</b> 를 위한 AI기술 개발 프로젝트 <b>Aria</b>
8	AI 안전성 및 윤리	- 책임 있는 AI 개발 원칙과 접근 방식 소개 Reasonable AI at Meta

\* 주요 논문성과, 뉴스기사, 주요인사의 인터뷰 기반으로 정리

- 메모리 연구에서는 end-to-end Memory Networks와 Gradient Episodic Memory를 개발하여 외부메모리 활용과 지속적 학습을 가능케 하고 있으며, 불안정한 정보 게임을 위한 심층 강화학습 ReBeL을 통해 계획 및 추론 능력을 향상
- OK-Robot을 통해 비전언어모델과 이동제어, 객체조작 등 로봇 기반 요소기술을 통합하고, MODeL을 통해 텐서수명과 메모리 위치 최적화를 실현
- 실세계 응용 측면에서는 메타버스 비전을 위한 AI 핵심 요소기술 개발과 AR 글래스를 위한 Aria 프로젝트를 추진하고 있음

35) 메타, 올해 약 35만 개 GPU 보유 계획 발표 (2024.1.22.) (출처 : <https://www.digitaltoday.co.kr/news/articleView.html?dxno=502824>)

- ‘Reasonable AI at Meta’ 정책을 통해 책임 있는 AI 개발 원칙과 전략을 공개하며, 안전하고 윤리적인 AGI 개발을 위한 방향성을 제시
- (주요 연구 분야) 컴퓨터과학 전반을 포괄하는 학회에서의 연구활동이 가장 많으며, 구글 스칼라에서 제시하는 H5-index 기준 인공지능과 컴퓨터비전 및 패턴인식 분야의 상위 학회들에서도 논문활동 활발함

**[주요 연구 분야 - 메타]**

학회 및 저널명	주요 세부 영역	논문수
Lecture Notes in Computer Science	인공지능, 데이터베이스, 소프트웨어 공학, 알고리즘 이론, 컴퓨터 네트워크, 정보 보안	290
Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition	컴퓨터비전, 패턴인식, 3D 컴퓨터 비전, 영상 생성 및 변환	251
Advances in Neural Information Processing Systems	딥러닝 및 머신러닝, 강화학습, 신경망아키텍처, 인공지능 이론	234
Proceedings of Machine Learning Research	기계학습, 통계적학습이론, 비지도 학습 및 지도학습, 인공지능 및 로봇틱스	145
International Conference on Learning Representations	자기지도학습, 생성 모델, 강화학습, 신경망구조	138

\* 2013년 이후, 총 4,923건

- (주요 협력관계) 상위 10개 기관 중 8개 곳이 자국내 위치한 기관이며, 대부분 대학과 협력 중

**[주요 협력관계 - 메타]**

협력 기관명	협력 논문수	비고
Carnegie Mellon University	246	미국
Google LLC	232	미국, 기업
Stanford University	206	미국
University of California, Berkeley	188	미국
New York University	182	미국
Georgia Institute of Technology	157	미국
The University of Texas at Austin	151	미국
University College London	143	영국
Tel Aviv University	131	이스라엘
Cornell University	99	미국

- (주요 국내 협력관계) 서울대, 연세대, KAIST, 고려대, UNIST 등 대학들과 주로 협력

[국내 기관과의 협력관계 - 메타]

협력 기관명	협력 논문수	비고
서울대학교	20	대학
연세대학교	14	대학
KAIST	11	대학
고려대학교	5	대학
UNIST	5	대학
ETRI	1	연구기관

\* 2013년 이후, 총 68건

□ 앤스로픽

- 스케일링 가설에 기반한 모델 성능 극대화를 추구하면서도, AI 시스템의 안전성과 윤리성을 핵심 가치로 삼아 책임있는 AGI 개발을 위해 노력하고 있으며, 모델의 해석 가능성 향상과 환각 현상 감소 등 AI 안전 관련 연구에서 성과 도출
  - Claude 시리즈를 통해 대규모언어모델을 개발하면서, 텍스트와 이미지를 함께 처리할 수 있는 멀티모달 기능을 통합

[AGI 연구 성과 및 방향 - 앤스로픽]

	주요 연구	연구 성과 및 방향
1	대규모언어모델	- 대규모언어모델 Claude 시리즈
2	멀티모달 학습	- 텍스트, 이미지를 함께 입력받는 Claude
3	계획 및 추론	- 인력보강으로 계획 및 추론 역량 강화
4	메모리	- 대화 내용 기억 prompt caching
5	에이전트	- AI모델이 컴퓨터를 직접 조작하는 Computer Use - AI모델과 다양한 데이터소스와 연결을 지원한 도구 MCP
6	실세계 문제해결	- 대화형 인공지능 챗봇 Claude
7	AI 안전성 및 윤리	- AI 시스템에 특정 원칙과 가치를 내재화, Constitutional AI - 대규모언어모델 작동방식 해석 - 환각현상을 50% 감소 Claude 2.1

\* 주요 논문성과, 및 뉴스기사, 주요인사의 인터뷰 기반으로 정리

\* 오픈AI에서 퇴사한 연구원들이 2021년 설립한 앤스로픽은 최근 오픈AI 공동 창립자이자 강화학습에서 뛰어난 연구 업적을 갖춘 존 솔만이 합류함으로써(2024.8) AGI 개발에서 계획 및 추론 역량이 한층 강화됨

- 인력보강을 통한 계획 및 추론 역량 강화와 대화 내용 기억을 위한 prompt caching 기술을 개발하여 모델의 성능을 지속적으로 향상 중
- 최근에 공개된 Computer Use 기능을 통해 AI 모델이 컴퓨터를 직접 조작할 수 있게 하고, Claude를 통해 고도화된 대화형 인공지능을 구현

- 특히 앤스로픽은 여타 AGI 개발 기업보다도 AI 헌법을 강조하며 안전한 AGI 개발을 추구
  - Constitutional AI를 통해 AI 시스템에 특정 원칙과 가치를 내재화하고, 대규모언어모델의 작동 방식을 해석하며, Claude 2.1에서 환각 현상을 50% 감소시키는 등 AI 안전성 확보에 주력
- (주요 연구 분야) 논문활동은 2022년부터 이루어져 총 논문 수는 21편으로 많지 않은 편, 안전을 최우선으로 생각하는 기업답게 AI 안전 관련 논문을 2024년에 게재

**[주요 연구 분야 - 앤스로픽]**

학회 및 저널명	주요 세부 영역	논문수
Proceedings of Machine Learning Research	기계학습, 통계적학습이론, 비지도 학습 및 지도학습, 인공지능 및 로봇틱스	10
Advances in Neural Information Processing Systems	딥러닝 및 머신러닝, 강화학습, 신경망아키텍처, 인공지능 이론	2
Proceedings of the Annual Meeting of the Association for Computational Linguistics	기계 번역, 정보 추출, 감성 분석, 대화 시스템, 언어 모델링, 텍스트 요약	2
ACM Conference on Fairness, Accountability, and Transparency	알고리즘 편향성, AI 시스템의 공정성, 기계학습 모델의 해석 가능성, 데이터 프라이버시, AI의 사회적, 법적 영향	1

\* 2013년 이후, 총 21건

- (주요 협력관계) 상위 10개의 기관 중 9개의 기관이 자국의 기관이며, 대학과 다수 협력 중이나 기업인 메타, MS, Cohere와도 협력

**[주요 협력관계 - 앤스로픽]**

협력 기관명	협력 논문수	비고
Meta	7	미국, 기업
Massachusetts Institute of Technology	5	미국
New York University	4	미국
Cornell University	3	미국
University of Southern California	2	미국
Microsoft Corporation	2	미국, 기업
Cohere	2	캐나다, 기업
Harvard University	2	미국
Cornell University	2	미국

- (주요 국내 협력관계) 국내 기관과의 협력관계는 존재하지 않음

□ AGI 연구 방향

○ 대규모언어모델을 기반으로 멀티모달리티를 강화하고 추론능력을 향상

- 기본모델은 트랜스포머 아키텍처 기반 대규모멀티모달모델을 개선하는 방향으로 주로 연구되고 있으나 트랜스포머를 넘어서는 새로운 아키텍처 연구도 병행
- 추론능력 향상을 위하여 뉴럴-심볼릭AI, 강화학습, 생각의 사슬(Chain of Thought), 인과 추론(Causal Reasoning), 메타학습(Meta Learning) 등 연구 진행

[AGI 세부기술별 빅테크 기업들의 연구동향 비교]

세부 기술	주요 연구	구글	답마인드	오픈AI	메타	앤스로픽
AGI 기본모델	대규모언어모델	○	○	○	○	○
	멀티모달 학습	○	○	○	○	○
	계획 및 추론	○	○	○	○	○
	메모리	○	○	○	○	○
	메타학습	○	○	○		
	신경과학과 AI		○			
	오픈엔드 시스템		○			
인터페이스	에이전트	○	○	○		○
시스템 효율화	자동화된 기계학습	○			○	
안전과 신뢰	AI 안정성 및 윤리	○	○	○	○	○
AGI 응용	로봇공학	○	○	○	○	
	실세계 문제해결	○	○	○	○	○
	양자컴퓨터와 AI	○				
	AI 칩 설계 및 개발	○		○		

\* 해당 기관들의 비공개 연구 동향 및 분석 대상 문서에 포함되지 않은 기술들은 위 표에 표기되지 않음

- 최근 대규모언어모델의 성능 향상 한계가 언급되면서 추론을 비롯한 사후훈련과 강화학습의 중요성이 강조되고 있음

\* 오픈AI와 구글은 사전훈련 중 모델의 학습방법을 설정하는 하이퍼파라미터 튜닝을 비롯해 기본모델의 미세조정을 더 정교하게 하고 인간피드백을 강화하거나 기존 대규모모델을 활용해 고품질 합성데이터를 생성하여 활용하거나 모델에 같은 질문을 수십, 수백번 한 다음 가장 좋은 답을 선택하는 샘플링을 활용하는 등 추론 강화를 위해 노력 중<sup>36)</sup>

○ 추론 성능이 향상된 AI 모델들이 2024년 하반기 대거 출시되면서 디지털기기들을 작동하며 복잡한 작업을 수행할 수 있는 에이전트 연구도 활발

- 컨텍스트 윈도우 (단기 기억) 와 모델 가중치(장기 기억) 사이의 간극을 메울 수 있는 일화 메모리(Episodic Memory)를 개발하여 개인서비스 강화

36) 한계달한 스케일링법칙...AGI 꿈 멀어지나 (2024.11.16.) (출처: <https://www.aitimes.com/news/articleView.html?idxno=165343>)

- 경제성과 환경을 고려한 학습 알고리즘을 개발하기 위해 노력
  - 첨단 모델로부터 고급학습 데이터를 생성하여 작지만 성능은 강력한 소형모델 개발
- 로봇지능으로 활용하기 위한 물리적 세계 이해 모델에 대한 기초 연구 진행

## 5 국내 기업의 AGI 개발 동향

- 대형언어모델의 파운데이션모델을 자체적으로 개발한 기업 중심으로 기관 홈페이지 및 뉴스 보도를 기반으로 정리
  - 글로벌협력은 파운데이션모델과 관련된 글로벌 협력 위주로 조사

### □ 네이버 AI 랩

- ChatGPT에 비해 한국어를 6,500배 더 많이 학습하여 자연스러운 한국어 표현은 물론, 한국 사회의 법, 제도, 문화적 맥락까지 이해하며 소통 가능한 대형언어모델 **하이퍼클로바X** 최신 버전을 2024년 7월에 공개
  - \* 2021년 공개된 모델의 매개변수 수는 204B개, 2023년 공개된 하이퍼클로바X는 300B ~ 400B개로 추정, 2024년 7월 공개된 모델은 82B개
- 한국어와 영어뿐만 아니라 일본어, 중국어 등도 처리
  - 텍스트, 이미지, 음성 등 다양한 데이터 처리를 위한 멀티모달 능력 확보
- 우수한 LLM 개발 역량을 기반으로 사우디와 협력해 아랍어 소버린AI를 구축하기 위한 협의 진행 중<sup>37)</sup>

### □ SK텔레콤

- 인공지능 기반의 개인화된 음성비서 에이전트 **A.(에이닷)** 개발
- 자연어처리, 멀티모달 인터페이스, 개인화, IoT 연동 등에 AI를 활용해 편리한 일상생활 서비스 지원을 위해 노력
- 오픈AI의 GPT-4o, 앤스로픽의 Claude 3.5 소네트, 퍼플렉시티의 대규모언어모델 등을 자체개발 언어모델과 함께 서비스

### □ KT

- 기업고객을 대상으로 서비스하기 위한 맞춤형 대규모언어모델 AI **민:음** 개발
- 민:음 모델과 3자 모델들을 통합하여 B2B 특화 LLM 모델 개발

37) 네이버, 韓 AI 대표주자 발돋움...사우디와 LLM 협업 임박 (2024.9.11.) (출처: <https://m.edaily.co.kr/News/Read?newsId=03788406639019136&mediaCodeNo=257>)

- KT는 MS와 인공지능 협력에 5년 동안 2조 4000억 원을 투자하고 한국어 특화 AI 모델과 서비스 개발, 한국형 클라우드 서비스 개발, AX 전문기업 설립 등을 추진하기로 하였으며, 한국어 특화 AI 모델은 MS가 최대 주주로 있는 오픈AI의 GPT-4o 기반으로 개발해 2025년 2분기에 출시할 계획<sup>38)</sup>

□ **LG AI 연구원**

- 2024년 8월 대형언어모델 **EXAONE 3.0**을 발표하였으며, 모델라인업 중에서 7.8B Instruction Tuned 모델을 연구용으로 활용할 수 있도록 오픈소스로도 공개
- 7.8B 모델과 유사한 크기의 모델들, Llama3.1 8B Inst., Gemma2 9B Inst., Phi3 7B Inst., Mistral 7B Inst., 등과 성능비교 결과, Math Coding의 경우 평균 점수 1위로 타 모델대비 우위를 보였고, 추론(reasoning)에서도 높은 성능을 달성
- EXAONE을 텍스트, 음성, 이미지, 비디오 등 모든 것이 결합된 멀티모달모델로 확장하고, 스스로 데이터를 수집·가공하고 답을 하는 에이전트를 개발하는데 주력<sup>39)</sup>

□ **삼성**

- 삼성은 2023년 초 대형언어모델 개발에 착수하여 그해 말 **가우스** 시리즈를 발표
  - \* 언어모델 Gauss Language, 이미지모델 Gauss Image, 코딩모델 Gauss Code
- 보안에 강한 온디바이스 AI와 고성능 클라우드 AI를 결합해 최적화하고, 개인정보보호와 보안 기술 연구에 집중
- 미세조정 작업 중 모델의 성능향상에 필요한 매개변수의 변경사항만을 학습하는 LoRA를 도입하여 대규모언어모델의 가드레일을 효율적으로 적용할 수 있는 기술 개발
- Galaxy AI 기기에 자체 개발한 가우스와 구글의 제미나이 나노가 함께 탑재되었으며, 오픈AI와도 유사 협력을 협의 중<sup>40)41)</sup>

□ **포티투마루**

- 2024년 5월 기업용 초거대 AI 경량화 모델인 **LLM42** 출시
- 대형언어모델의 대표적 단점인 환각현상을 검색증강생성 기술인 RAG와 인공지능 독해기술 MRC(Machine Reading Comprehension)을 적용해 완화

38) KT, MS와 협력... 내년 상반기 '한국형 AI' 출시 (2024.10.11.) (출처: <https://www.donga.com/news/Economy/article/all/20241010/130193502/2>)

39) LG AI 연구원, 6개월마다 생성형AI 엑사원 신모델 내놓는다 (2024.10.24.) (출처: <https://www.newspim.com/news/view/20241024001066>)

40) 삼성 갤럭시 S24 시리즈에 탑재된 구글 AI (2024.1.17.) (출처: <https://blog.google/intl/ko-kr/products/android-play-hardware/google-ai-samsung-galaxy-s24/>)

41) 오픈AI, 삼성갤럭시에 AI 탑재논의 (2024.11.22.) (출처: <https://www.etnews.com/20241122000290>)

\* LLM은 언어 전반에 대한 이해와 생성 능력을 갖춘 범용모델인 반면, MRC는 주어진 지문을 이해하여 질문에 답하는데 특화됨

- 글로벌 대표 앤서링 AI 기술력을 인정받으며 2021년 시리즈 AI 투자유치에 이어 2024년 초 LG유플러스 100억 원, 한컴 40억 원 등 시리즈 B 투자유치
- 최근 자사 대규모언어모델에 대하여 TTA에서 운영하는 AI 신뢰성 인증 획득<sup>42)</sup>
- 글로벌 뉴스 제공 업체 로이터(Thomson Reuters Corporation Pte Limited)와 생성형 AI 모델을 활용한 새로운 미디어 서비스 구축을 위한 전략적 파트너십 체결<sup>43)</sup>

□ **업스테이지**

- 2023년 12월 경량모델이지만 기존 대형모델들의 성능을 능가하는 7B **솔라(Solar)** 공개
- 허깅페이스에서 운영하는 오픈 LLM 리더보드에서 세계 1위 달성하였으며 Mistral 8x7B 모델과 비교하여도 더 뛰어난 성과를 보여줌(2023년 12월 평가 기준)
- 대형언어모델 구축에 특화된 광학문자인식(OCR) 기술을 활용하여 데이터 학습 및 인식 속도를 극대화

□ **코난테크놀러지**

- 자체 구축한 양질의 대규모 데이터, 컴퓨팅 인프라, 20년 이상 축적한 자연어처리기술을 활용하여 2023년 8월 47B 대규모언어모델 **코난LLM**을 출시
- 2024년 7월 온디바이스용 4B와 7B 모델을 갤럭시 24에 탑재하여 STT(Speech to Text), TTS(Text to Speach), LLM 기능 제공
  - \* 사람중심인공지능핵심원천기술개발사업의 ‘플러그앤플레이 방식으로 설명 가능성을 제공하는 인공지능 기술 개발 및 인공지능 시스템에 대한 설명 제공 검증’ 과제의 공동연구기관으로 참여

□ **트릴리온랩스**

- 네이버 하이퍼클로버X의 핵심개발 연구원과 국내외 대학에서 LLM 개발을 연구한 경험을 보유하고 있는 인재들이 설립한 스타트업
- **한국어 및 한국 문화에 가장 특화된 LLM 파운데이션 모델을 2024년 말 출시할 계획을 발표**
- 2024년 9월 420만 달러(약 57억 원) 규모의 프리시드 투자 유치

42) 포티투마루, TTA AI 신뢰성 인증 획득 (2024.11.6.) (출처: <http://www.itdaily.kr/news/articleView.html?idxno=228481>)

43) 포티투마루, 글로벌 통신사 로이터와 전략적 파트너십 체결 (2024.10.18.) (출처: <https://www.aitimes.kr/news/articleView.html?idxno=32474>)



## 6 소결

### □ AGI 개발 현 수준

- 기존의 모호했던 **AGI 개념과 기술수준 진단**이 빅테크 기업들의 단계별 정의를 통해 **구체화**
  - 단일 시점의 획기적 결과물이 아닌 다양한 능력의 점진적 통합과 실제 응용을 통한 발전 과정으로 인식 전환되며, 現 기술 수준을 평가하고 향후 기술 로드맵을 수립을 위한 실제적인 논의 시작
  - 현재의 AI는 대부분 딥마인드가 정의한 AGI 1단계 수준이며, 일부 오픈AI가 정의한 AGI 2단계 수준의 초입에 진입
- **대형언어모델을 중심으로** 멀티모달능력을 확장하고 추론 능력을 강화하는 방향으로 발전
  - 개념적 수준에 머물러 있던 AGI 연구는 최근 대규모언어모델의 등장으로 동력 마련
  - 구글, 딥마인드, 오픈AI 등은 대규모 컴퓨팅 파워를 활용한 대형모델 개발에 주력
  - 메타는 오픈소스 전략으로 협력적 발전을 추구
  - 앤스로픽은 AI 안전성을 최우선 고려하며 개발
- 트랜스포머를 넘어선 AGI 개발을 위한 **새로운 아키텍처 개발** 연구도 진행
  - 딥마인드는 오픈엔드 학습을 AGI의 본질로 보고 관련 연구 강화
  - 신경과학 기반의 뉴로-심볼릭 AI 시스템 연구
- **추론 능력 향상**과 메모리 기능 등 개인화 서비스가 가능한 기술들이 갖추어짐에 따라 최근 디지털기기들을 작동하며 복잡한 작업을 수행할 수 있는 **에이전트 개발 경쟁 심화**
- 국내 기업들은 한국어에 특화된 자체 대형언어모델 개발에 주력
  - 네이버, LG AI 등 대기업은 자체 대형언어모델 개발
  - 포터투마루, 업스테이지 등 AI 스타트업은 특화 기술로 경쟁력 확보

### □ 향후 과제

- AGI 실현을 위한 핵심 기술 확보
  - 자율성장, 추론능력, 메모리 관리 등 AGI의 핵심 기능 개발
  - 멀티모달 처리 능력과 실세계 상호작용 능력 강화
  - 효율적인 학습 알고리즘과 컴퓨팅 자원 최적화 기술 확보
- AGI 안전성 및 신뢰성 확보
  - 설명 가능성, 편향성 제거, 윤리적 의사결정, 프라이버시 보호와 데이터보안 강화

- 인간 가치와의 정렬(alignment) 방안 연구
- o 차별화된 전략적 접근 필요
  - 한국어 특화 및 문화적 맥락 이해 강화
  - 제한된 컴퓨팅 자원 하에서의 효율적 모델 개발
  - 사용자 니즈를 예측하여 틈새 영역에서의 전문가 수준 AGI 개발 전략 수립
  - 산학연 협력 강화 및 글로벌 연구 네트워크 확대

## 참고문헌

2023 Expert Survey on Progress in AI, AI Impact, 2023.10.

(출처: <https://blog.aiimpacts.org/p/2023-ai-survey-of-2778-six-things>)

OpenAI's Sam Altman says human-level AI is coming but will change world much less than we think, CNBC (2024.1.16.)

(출처: <https://www.cnn.com/2024/01/16/openai-sam-altman-agi-coming-but-is-less-impactful-than-we-think.html>)

Shane Legg (DeepMind Founder) - 2028 AGI, Superhuman Alignment, New Architectures (2023.11)

(출처: <https://www.youtube.com/watch?v=Kc1atfJkiJU>)

Tesla's Musk predicts AI will be smarter than the smartest human next year, Reuters (2024.4.9.)

(출처: <https://www.reuters.com/technology/teslas-musk-predicts-ai-will-be-smarter-than-smartest-human-next-year-2024-04-08/>)

OpenAI looks to shift away from nonprofit roots and convert itself to for-profit company, AP (2024.9.26.)

(출처: <https://apnews.com/article/chatgpt-openai-sam-altman-nonprofit-859bff5c19845f51796244e0072e2dfb>)

State of AI 2023 Report, CBInsight (2024.2.1.)

(출처: <https://www.cbinsights.com/research/report/ai-trends-2023/>)

Mark Gubrud. Nanotechnology and International Security. Fifth Foresight Conference on Molecular Nanotechnology, November 1997.

Levels of AGI: Operationalizing Progress on the Path to AGI, Google DeepMind, arXiv:2311.02462v2 (2024.1)

OpenAI's 5 Levels of Super AI (AGI To Outperform Human Capability), Forbes (2024.7)

(출처: <https://www.forbes.com/sites/jodiecook/2024/07/16/openai-5-levels-of-super-ai-agi-to-outperform-human-capability/>)

汎用AIの現状と展望 - 実現への見通しと社会への影響, JRI (2024.6)

What is artificial general intelligence (AGI)?, McKinsey & Company (2024.03)

Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models, arXiv:2206.04615 (2022.9)

Think you have Solved Question Answering? Try ARC, the AI2 Reasoning Challenge, arXiv:1803.05457 (2018.4)

BBQ: A Hand-Built Bias Benchmark for Question Answering, arXiv:2110.08193 (2021.10)

RobustBench: a standardized adversarial robustness benchmark, arXiv:2010.09670 (2020.10)

SafeLife, Github (출처: <https://github.com/PartnershipOnAI/safelife>)

FrontierMath, EpochAI (출처: <https://epoch.ai/frontiermath>)

MLE-bench (2024.10.10.) (출처: <https://openai.com/index/mle-bench/>)

LoTa-Bench: Benchmarking Language-oriented Task Planners for Embodied Agents, arXiv:2402.08178 (2024.2.13.)

How Far Are We From AGI?, arXiv:2405.10313v1 (2025.5.16.)

Artificial Intelligence Index Report, HAI(Human-Centered Artificial Intelligence) (2024.4)

Introducing OpenAI o1-preview, OpenAI (2024.9.12.)

(출처: <https://openai.com/index/introducing-openai-o1-preview/>)

구글, AI 에이전트 ‘자비스’ 12월 공개 예정, AI타임즈 (2024.10.27.)

(출처: <https://www.aitimes.com/news/articleView.html?idxno=164658>)

구글의 최첨단 양자 칩 ‘윌로우(Willow)’를 소개합니다, 구글코리아블로그 (2024.12.9.)

(출처: <https://blog.google/intl/ko-kr/company-news/technology/quantum-ai-willow-kr/>)

Open-Endedness is Essential for Artificial Superhuman Intelligenc, arXiv:2406.04268 (2024.6.6.)

ChatGPT 5: Everything we know so far about Orion, OpenAI’s next big LLM, TheIndianEXPRESS (2024.10.29.)

(출처: <https://indianexpress.com/article/technology/artificial-intelligence/openai-chatgpt-5-orion-llm-features-capabilities-launch-9641365/>)

OpenAI confirms AI agents are coming next year - what it means for you, Tom’s Guide (2024.10.2.)

(출처: <https://www.tomsguide.com/ai/chatgpt/the-agents-are-coming-openai-confirms-ai-will-work-without-humans-in-2025>)

OpenAI builds first chip with Broadcom and TSMC, scales back foundry ambition, Reuters (2024.10.31.)

(출처: <https://www.reuters.com/technology/artificial-intelligence/openai-builds-first-chip-with-broadcom-tsmc-scales-back-foundry-ambition-2024-10-29/>)

메타, 올해 약 35만개 GPU 보유 계획 발표, 디지털투데이 (2024.1.22.)

(출처 : <https://www.digitaltoday.co.kr/news/articleView.html?idxno=502824>)

한계달한 스케일링법칙...AGI 꿈 멀어지나, AI타임즈 (2024.11.16.)

(출처: <https://www.aitimes.com/news/articleView.html?idxno=165343>)

네이버, 韓 AI 대표주자 발돋움...사우디와 LLM 협업 임박, 이데일리 (2024.9.11.)

(출처: <https://m.edaily.co.kr/News/Read?newsId=03788406639019136&mediaCodeNo=257>)

KT, MS와 협력... 내년 상반기 ‘한국형 AI’ 출시, 동아일보 (2024.10.11.)

(출처: <https://www.donga.com/news/Economy/article/all/20241010/130193502/2>)

LG AI 연구원, 6개월마다 생성형AI 엑사원 신모델 내놓는다, 뉴스핌 (2024.10.24.)

(출처: <https://www.newspim.com/news/view/20241024001066>)

삼성 갤럭시 S24 시리즈에 탑재된 구글 AI, 구글코리아블로그 (2024.1.17.)

(출처: <https://blog.google/intl/ko-kr/products/android-play-hardware/google-ai-samsung-galaxy-s24/>)

오픈AI, 삼성갤럭시에 AI 탑재논의, 전자신문 (2024.11.22.)

(출처: <https://www.etnews.com/20241122000290>)

포티투마루, TTA AI 신뢰성 인증 획득, IT데일리 (2024.11.6.)

(출처: <http://www.itdaily.kr/news/articleView.html?idxno=228481>)

포티투마루, 글로벌 통신사 로이터와 전략적 파트너십 체결, AI타임즈 (2024.10.18.)

(출처: <https://www.aitimes.kr/news/articleView.html?idxno=32474>)

---

저자 소개

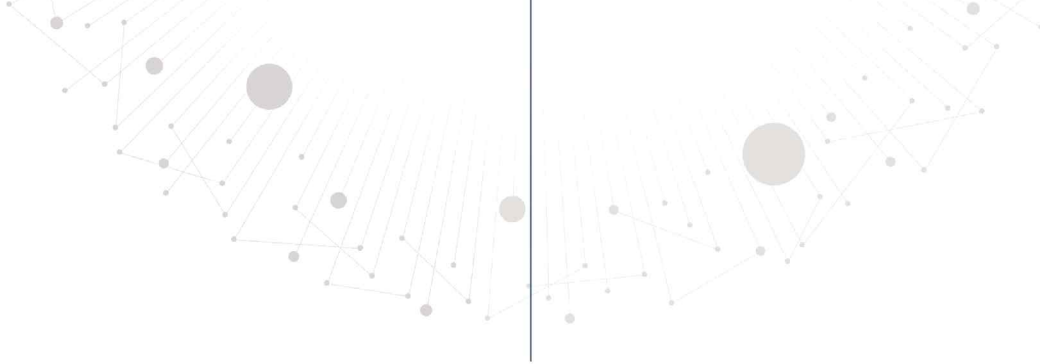
정선화 ETRI ICT전략연구소 기술정책연구본부 기술경제연구실 책임연구원  
e-mail: sh-jeong@etri.re.kr Tel. 042-860-6511

---

AGI 기술개발 동향 - 주요 빅테크 기업 중심으로

발행인 한 성 수  
발행처 한국전자통신연구원 ICT전략연구소  
발행일 2024년 12월 31일





[www.etri.re.kr](http://www.etri.re.kr)

본 저작물은 공공누리 제4유형:  
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.



**ETRI** Electronics and Telecommunications  
Research Institute

34129 대전광역시 유성구 가정로 218  
TEL.(042) 860-6114 FAX.(042) 860-6504

