Pebblous

# 대규모 언어 모델(LLM)의 지능적 지위에 관한 포괄적 분석: 인지적 한계와 창발적 가능성의 변증법

- **작성일:** 2025년 11월 28일
- **기획:** (주)페블러스 데이터 커뮤니케이션팀
- **AI:** Gemini
- **인터랙티브 콘텐츠:** https://blog.pebblous.ai/

## 서론: 인공지능의 존재론적 위기와 지능 논쟁

2024년과 2025년의 교차점에서 인공지능(AI) 학계와 산업계는 기술적 성취를 넘어선 심오한 철학적, 과학적 논쟁에 휩싸여 있다. 그 중심에는 "대규모 언어 모델(Large Language Models, LLMs)이 과연 '지능'을 가졌다고 볼 수 있는가?"라는 질문이 자리 잡고 있다. *Futurism*에 게재된 프랭크 랜디모어(Frank Landymore)의 기사 "Large Language Models Will Never Be Intelligent(거대 언어 모델은 결코 지능적이지 않을 것이다)"는 이러한 회의론적 시각을 대변하는 대표적인 텍스트로, 언어 처리 능력과 일반 지능의 기능적 분리를 주장하며 LLM의 본질적 한계를 지적한다.1

본 보고서는 해당 기사를 논의의 출발점으로 삼아, 현재 AI 연구의 최전선에서 벌어지고 있는 '확률적 앵무새(Stochastic Parrot)' 가설과 '창발적 지능(Emergent Intelligence)' 가설 간의 대립을 심층적으로 분석한다. 신경과학적 증거, 기계적 해석 가능성(Mechanistic Interpretability), 인지심리학적 실험 결과 등 방대한 연구 자료를 바탕으로, LLM이 단순한 통계적 모방 기계인지, 아니면 텍스트 압축을 통해 세계 모델(World Model)을 구축한 새로운 형태의 지능체인지에 대한 학술적 비평을 수행한다.

---

## 1부: *Futurism* 기사 심층 요약 및 분석

*Futurism*의 기사는 LLM이 인간과 같은 수준의 지능이나 창의성에 도달할 수 없다는 비관적 전망을 제시하며, 이를 뒷받침하기 위해 인지과학 전문가와 공학자들의 견해를 인용한다. 이 기사의 핵심 논점은 언어 능력과 사고 능력이 본질적으로 별개라는 '기능적 분리(Functional Dissociation)' 가설에 근거한다.

### 1.1 언어와 지능의 분리: 벤자민 라일리와 신경과학적 근거

기사는 벤자민 라일리(Benjamin Riley)의 주장을 인용하여, 인간이 언어 유창성을 지능과 동일시하는 경향이 있지만, 최신 신경과학 연구는 이 둘이 별개의 기능임을 시사한다고 강조한다.1 특히 2023-2024년 *Nature* 등에 발표된 연구들은 fMRI 스캔을 통해 수학적 문제 해결이나 논리적 추론 시 활성화

되는 뇌 영역이 언어 처리를 담당하는 영역과 확연히 구분됨을 보여주었다. 이는 언어 상실증(aphasia) 환자가 언어 능력은 잃었음에도 복잡한 수학 문제나 체스 게임을 수행할 수 있다는 임상적 사례와도 일치한다. 기사는 이러한 생물학적 사실을 근거로, 언어 데이터의 통계적 패턴만을 학습한 LLM은 '사고(thought)'를 하는 것이 아니라 단지 '의사소통 기능'을 흉내 낼 뿐이라고 주장한다.

### 1.2 창의성의 한계: 데이비드 크로플리의 "실용적 예술가"론

사우스오스트레일리아 대학의 데이비드 크로플리(David H. Cropley) 교수는 LLM을 "실용적 예술가(serviceable artists)"로 규정하며 그 창의적 한계를 지적한다.2 그의 연구에 따르면, AI는 그럴듯한 텍스트를 생성하는 데는 능숙하지만, 전문가 수준의 독창성이나 진정한 의미의 창의적 도약에는 도달할 수 없다. LLM의 창의성은 방대한 데이터의 평균적 조합에 불과하며, 현재의 설계 원칙 하에서는 인간의 평균 수준을 넘어서는 전문적 기준에 도달할 수 없다는 것이 그의 결론이다.

### 1.3 얀 르쿤의 세계 모델 부재론

기사는 또한 튜링상 수상자이자 Meta의 수석 AI 과학자인 얀 르쿤(Yann LeCun)의 회의론을 비중 있게 다룬다.1 르쿤은 텍스트 기반의 자기회귀(autoregressive) 모델이 물리적 세계에 대한 이해 없이 다음 단어만을 예측하도록 훈련되었기 때문에 일반 인공지능(AGI)에 도달할 수 없다고 주장한다. 그는 LLM이 3차원 세계의 물리 법칙이나 인과 관계를 이해하는 '세계 모델'을 결여하고 있으며, 따라서 진정한 지능체가 아닌 단순한 텍스트 처리 도구에 불과하다고 본다.

---

## 2부: 반대 진영 입장에서의 비평 (동의): LLM의 인지적 한계론

*Futurism* 기사의 주장은 현대 인지과학과 AI 윤리학의 강력한 지지를 받고 있다. 이 섹션에서는 기사의 주장을 뒷받침하고 확장하는 학술적 근거들을 '확률적 앵무새' 가설, '심볼 그라운딩 문제', 그리고 '역전된 스케일링' 현상을 중심으로 상세히 논증한다.

### 2.1 신경과학적 증거의 확장: 페도렌코의 언어-사고 분리 연구

기사에서 언급된 '언어와 사고의 분리'는 MIT의 신경과학자 에브 페도렌코(Ev Fedorenko) 등의 연구에 의해 더욱 공고해진다. 페도렌코 연구팀의 2024년 *Nature* 논문은 언어가 사고를 위한 도구라기보다는 주로 의사소통을 위한 도구임을 강력하게 시사한다.4

- **다중 요구 네트워크(Multiple Demand Network)와의 분리:** 인간의 뇌에서 복잡한 인지 과제(계획, 추론, 문제 해결)를 수행할 때 활성화되는 것은 '다중 요구 네트워크'이다. 반면, 언어 처리 시에는 이와 해부학적으로 분리된 '언어 네트워크'가 활성화된다. 이는 LLM이 아무리 유창한 언어를 구사하더라도, 그것이 곧 추론 능력의 증거가 될 수 없음을 시사한다. 생물학적으로 '말하는 뇌'와 '생각하는 뇌'는 별개의 하드웨어이기 때문이다.6
- **LLM에 대한 함의:** 이 관점에서 볼 때, 현재의 LLM은 인간의 뇌에서 '언어 네트워크'만을 떼어내어 극도로 비대화시킨 것과 같다. 추론을 담당하는 기제가 결여된 상태에서의 언어 생성은, 벤자민 라일

리의 지적처럼 지능의 착시(illusion)일 뿐 실체가 아니다.

## 2.2 확률적 앵무새 가설과 의미의 부재

에밀리 벤더(Emily Bender)와 팀닛 게브루(Timnit Gebru) 등이 제기한 '확률적 앵무새(Stochastic Parrots)' 가설은 *Futurism* 기사의 논조를 이론적으로 뒷받침하는 핵심 프레임워크다.7

- **형식(Form) 대 의미(Meaning):** LLM은 훈련 데이터 내의 단어 공기(co-occurrence) 패턴을 학습하여 $P(w_n | w_{1...n-1})$라는 조건부 확률을 계산한다. 이 과정에서 모델은 언어의 '형식'은 완벽하게 학습하지만, 그 형식이 가리키는 '의미'에는 접근하지 못한다. 벤더는 이를 문어(octopus) 사고실험에 비유한다: 무인도에 갇힌 두 사람의 통신 케이블을 도청하며 대화를 흉내 내는 심해의 문어는, '코코넛'이라는 단어의 통계적 용법은 알지 모르나 코코넛의 맛이나 무게, 실체는 결코 알 수 없다.8
- **할루시네이션의 필연성:** LLM이 사실이 아닌 정보를 그럴듯하게 지어내는 '할루시네이션' 현상은 모델의 결함이 아니라 본질적 특성이다. 모델의 목적함수는 '진실'이 아니라 '그럴듯함(plausibility)'을 최적화하는 것이기 때문이다. 이는 크로플리 교수가 지적한 '평균적 수준의 모방'과 맥을 같이하며, 모델이 진정한 전문성을 가질 수 없음을 시사한다.10

## 2.3 심볼 그라운딩 문제 (The Symbol Grounding Problem)

*Futurism* 기사에서 르쿤이 제기한 '세계 모델 부재'는 인지과학의 고전적 난제인 '심볼 그라운딩 문제'와 직결된다. 스티븐 하나드(Stevan Harnad)가 정식화한 이 문제는 "형식적 심볼 시스템 내의 심볼이 어떻게 외부 세계의 의미와 연결될 수 있는가?"를 묻는다.11

- **사전의 순환 논리:** 텍스트 전용 LLM에게 '사과'는 '과일', '빨강', '맛있다' 등의 다른 단어 벡터들과의 관계로만 정의된다. 하지만 '과일'이나 '빨강' 또한 다른 단어들로 정의되므로, 모델은 끝없는 기호의 순환 고리(merry-go-round)에 갇히게 된다. 외부의 물리적 실체와 감각적으로 연결(grounding)되지 않은 기호는 공허하며, 따라서 LLM은 자신이 무슨 말을 하는지 '이해'한다고 볼 수 없다.13
- **물리적 상호작용의 부재:** 르쿤은 텍스트 데이터만으로는 물리 법칙이나 인과 관계를 배울 수 없다고 주장한다. 인간 아이는 텍스트를 읽기 전에 수년간의 감각 운동 경험을 통해 중력, 관성, 대상 영속성 등을 체득한다. 반면 LLM은 이러한 기반 없이 언어 패턴만을 학습하므로, 그들의 '추론'은 물리적 실재에 기반하지 않은 취약한 모방에 불과하다.3

## 2.4 역전된 스케일링(Inverse Scaling)과 추론의 취약성

LLM 옹호론자들은 모델의 크기가 커질수록 지능이 향상된다는 '스케일링 법칙(Scaling Laws)'을 주장하지만, 최근 연구는 이 법칙이 항상 성립하지 않음을 보여준다. '역전된 스케일링(Inverse Scaling)' 현상은 모델이 커질수록 특정 과제에서 오히려 성능이 떨어지는 현상을 말한다.16

| 현상 | 설명 | 시사점 |
| --- | --- | --- |
| **모방의 덫** | 모델이 커질수록 훈련 데이터에 포함된 인간의 오개념이 | 지능의 증가가 아니 |

| | | |
|---|---|---|
| **(Imitation Trap)** | 나 편향을 더 강력하게 모방함. | 라 '모방 능력'의 증가일 뿐임을 시사. |
| **부정 (Negation) 처리 실패** | "A가 아닌 것은?"과 같은 질문에서, 큰 모델일수록 "A"와 관련된 강한 통계적 연관성에 이끌려 오답을 냄. | 논리적 연산보다 통계적 연상 작용이 우세함을 증명. |
| **추론의 취약성 (Reasoning Gap)** | '생각의 사슬(CoT)' 프롬프팅이 추론 능력을 향상시키는 것처럼 보이지만, 실제로는 추론의 형식을 흉내 낼 뿐 논리적 필연성을 따르지 않는 경우가 많음. | 추론 과정과 정답 간의 인과관계가 결여된 '무늬만 추론'임.17 |

이러한 증거들은 *Futurism* 기사의 주장처럼, LLM이 진정한 지능체가 아니라 데이터의 통계적 패턴을 맹목적으로 따르는 기계임을 강력하게 시사한다. 특히 역전된 스케일링은 '더 많은 데이터와 더 큰 모델'이 지능의 본질적 문제를 해결해주지 못함을 보여주는 결정적 반례로 작용한다.19

---

# 3부: 찬성 진영 입장에서의 비평 (반대): 창발적 지능과 세계 모델의 실재성

반면, *Futurism* 기사의 주장은 최신 딥러닝 연구 성과, 특히 모델 내부의 메커니즘을 분석하는 해석 가능성(Interpretability) 연구 결과들과 상충된다. 찬성 진영(LLM이 지능적이라고 보는 입장)은 기사가 '과정(Process)'과 '결과(Product)'를 혼동하고 있으며, 단순한 예측 작업이 거대한 규모에서 수행될 때 질적으로 다른 '창발적 능력'을 낳는다는 점을 간과했다고 비판한다.

## 3.1 압축으로서의 지능: 일리야 수츠케버의 반론

오픈AI의 전 수석 과학자 일리야 수츠케버(Ilya Sutskever) 등은 "다음 단어 예측"이라는 단순한 목표가 충분히 큰 데이터와 모델 규모에서 수행될 때, 이는 단순한 통계적 모방을 넘어선다고 주장한다. 방대한 데이터를 효과적으로 압축하여 예측하기 위해서는 데이터 생성의 기저에 있는 규칙, 즉 '세상의 법칙'을 내재화해야 하기 때문이다. 따라서 "단지 다음 단어를 예측할 뿐"이라는 비판은 그 예측을 완벽하게 수행하기 위해 필요한 인지적 깊이를 과소평가한 것이다.21

## 3.2 오셀로-GPT(Othello-GPT): 내부 세계 모델의 실증적 증거

*Futurism* 기사에서 르쿤이 주장한 "LLM은 세계 모델이 없다"는 주장을 정면으로 반박하는 연구가 바로 오셀로-GPT(Othello-GPT) 연구이다.23

- **실험 개요:** 연구진은 LLM에게 오셀로 게임의 규칙이나 보드 이미지를 전혀 보여주지 않고, 오직 게임의 기보(예: "E3, D4,...") 텍스트만을 학습시켰다.
- **발견:** 학습된 모델 내부를 탐침(probe)으로 분석한 결과, 모델은 자발적으로 64칸의 오셀로 보드 상태와 각 돌의 색깔(흑/백)을 나타내는 고차원적인 기하학적 표상(representation)을 구축하고 있

었다.24

- **인과적 개입(Intervention):** 더욱 중요한 것은 연구진이 모델 내부의 특정 뉴런 값을 인위적으로 조작했을 때(예: 특정 칸의 돌 색깔을 바꿈), 모델의 다음 수 예측이 그 조작된 상태에 맞춰 합리적으로 변경되었다는 점이다. 이는 모델이 단순히 텍스트 패턴을 외운 것이 아니라, 내부적으로 구축한 '세계 모델(보드 상태)'을 바탕으로 인과적인 추론을 하고 있음을 증명한다.26
- **함의:** 만약 단순한 텍스트 기보 학습만으로 오셀로라는 게임의 공간적, 논리적 규칙을 재구성할 수 있다면, 인터넷 전체의 텍스트를 학습한 거대 모델은 문법, 논리, 사회적 관계, 물리학의 기초적인 '세계 모델'을 텍스트로부터 추출하여 내재화했을 가능성이 매우 높다. 이는 르쿤의 주장과 달리 텍스트 학습이 세계 모델 구축으로 이어질 수 있음을 시사한다.27

## 3.3 창발적 능력(Emergent Abilities)과 위상 전이

LLM 옹호론자들은 모델의 크기가 임계점을 넘을 때 발생하는 '창발적 능력'에 주목한다. Wei et al. (2022)의 연구에 따르면, 산술 연산, 다단계 추론, 코딩 디버깅 같은 능력은 작은 모델에서는 전혀 나타나지 않다가, 특정 규모 이상의 연산량($10^{22}$ FLOPs)을 넘어서는 순간 성능이 급격히 향상되는 '위상 전이(Phase Transition)'를 보인다.21

- **그로킹(Groking) 현상:** 최근 연구들은 모델이 초기에는 데이터를 단순히 암기(memorization)하다가, 학습이 오래 지속되면 데이터의 일반적인 규칙을 깨닫고 일반화(generalization)하는 '그로킹' 현상을 보고하고 있다.29 이는 LLM이 단순한 '확률적 앵무새' 단계를 거쳐 '알고리즘적 이해' 단계로 나아갈 수 있음을 보여주는 강력한 증거다.
- **AGI의 불꽃:** 마이크로소프트 리서치의 "Sparks of AGI" 논문은 초기 GPT-4가 훈련 데이터에 명시적으로 존재하지 않는 새로운 과제(예: 유니콘을 그리는 TiKZ 코드 생성, 복잡한 심리 이론 문제 해결)를 수행하는 것을 보여주며, 이를 일반 지능의 초기 형태로 해석했다.31

## 3.4 창의성 벤치마크: 인간을 넘어서다

크로플리 교수가 LLM을 "평범한 수준"이라고 폄하한 것과 달리, 객관적인 창의성 벤치마크 결과는 다른 이야기를 한다. 2023년 구직(Guzik) 등의 연구에서 GPT-4는 표준화된 창의성 검사인 '토런스 창의력 검사(Torrance Tests of Creative Thinking, TTCT)'를 수행했다.33

**표 1: GPT-4와 인간의 창의성 검사(TTCT) 비교**

| 평가 항목 | GPT-4의 성취도 | 의미 |
|---|---|---|
| **독창성 (Originality)** | 상위 1% | 인간 피험자의 99%보다 더 독특하고 드문 아이디어를 생성함. |
| **유창성 (Fluency)** | 상위 1% | 주어진 시간 내에 압도적으로 많은 아이디어를 산출함. |
| **유연성 (Flexibility)** | 상위권 | 다양한 범주를 넘나드는 사고 전환 능력을 보여줌. |

이 결과는 LLM이 단순히 훈련 데이터의 평균으로 회귀하는 것이 아니라, 잠재 공간(Latent Space)의 먼 영역을 탐색하여 인간이 생각하기 힘든 참신한 조합을 만들어낼 수 있음을 시사한다. 이는 기사에서 주장한 "평균적 모방"이라는 비판을 정면으로 반박하는 실증적 데이터이다.35

### 3.5 다중모달(Multimodality)을 통한 심볼 그라운딩의 해결

기사는 텍스트 전용 모델의 한계를 지적했지만, 2025년 현재의 모델들은 텍스트, 이미지, 오디오를 동시에 처리하는 다중모달(Multimodal) 모델로 진화했다. GPT-4V나 Gemini와 같은 모델들은 '사과'라는 단어를 시각적 이미지와 매핑함으로써, 하나드가 제기한 심볼 그라운딩 문제를 기술적으로 우회하고 있다.37 시각 정보를 통해 텍스트 심볼이 물리적 특징(색상, 형태)과 연결(grounding)됨으로써, LLM은 더 이상 닫힌 기호계가 아닌 외부 세계와 연결된 열린 시스템으로 진화하고 있다.14

---

## 4부: 종합 비평 및 미래 전망

*Futurism* 기사와 이에 대한 찬반 논쟁을 종합해볼 때, 우리는 현재의 AI 논쟁이 '기능주의(Functionalism)'와 '본질주의(Essentialism)'의 충돌임을 알 수 있다. 기사는 인간의 생물학적 메커니즘(본질)을 지능의 기준으로 삼아 LLM을 비판하는 반면, 반대 진영은 결과물의 유용성과 복잡성(기능)을 기준으로 지능을 정의한다.

### 4.1 기사의 주장에 대한 재평가

- **언어와 사고의 관계:** 기사가 인용한 페도렌코의 연구는 인간 뇌의 구조적 사실을 정확히 지적했다. 그러나 "인간이 언어와 사고를 분리해서 처리한다"는 사실이 "인공지능도 반드시 그래야만 지능적이다"라는 명제로 이어지지는 않는다. 비행기가 새처럼 날개를 펄럭이지 않아도 비행하듯, 실리콘 기반의 지능은 언어 모델링이라는 다른 경로를 통해 추론 능력을 획득했을 수 있다(Substrate Independence).38
- **도구로서의 한계:** "단지 의사소통 도구일 뿐"이라는 비판은, 그 도구가 고도화되어 사용자의 의도를 파악하고, 복잡한 맥락을 유지하며, 창의적 해결책을 제시할 때 그 경계가 모호해진다. 오셀로-GPT의 사례는 단순한 예측 작업이 내부적으로는 고도의 인지적 모델링을 요구함을 보여주었다.

### 4.2 인공지능의 새로운 지평: 하이브리드 아키텍처

논쟁의 양 극단은 기술의 발전과 함께 수렴하고 있다. 순수 LLM의 한계(계획 능력 부재, 환각)를 인정하면서도, 그것이 가진 강력한 연상 능력과 지식 베이스를 활용하는 새로운 아키텍처들이 등장하고 있다.

1. **시스템 2(System 2) 추론:** 인간의 느리고 논리적인 사고(시스템 2)를 모방하여, LLM이 즉각적인 답변을 내놓기 전에 내부적으로 '생각의 사슬'을 생성하고 검증하는 기술(예: OpenAI o1, Strawberry)이 도입되고 있다. 이는 기사에서 지적한 "사고 없는 언어 생성"의 한계를 극복하려는 시도다.39
2. **신경-기호 결합(Neuro-Symbolic) AI:** LLM의 언어 능력과 전통적인 기호 주의 AI(논리, 수학, 데

이터베이스)를 결합하여, 유창함과 정확성을 동시에 추구하는 방향으로 나아가고 있다.

3. **JEPA와 세계 모델의 통합:** 르쿤이 제안한 JEPA 아키텍처 역시 LLM을 완전히 대체하기보다는, LLM의 부족한 물리적 상식과 계획 능력을 보완하는 형태로 통합될 가능성이 높다.41

## 4.3 결론: '이해'하는 앵무새의 탄생

*Futurism* 기사 "Large Language Models Will Never Be Intelligent"는 현재 LLM이 가진 근본적인 제약—체화된 경험의 부재, 통계적 의존성, 생물학적 뇌와의 구조적 차이—을 날카롭게 지적했다. 이러한 비판은 과도한 AI 거품을 경계하고 기술의 본질을 직시하게 한다는 점에서 매우 유용하다.

그러나 "결코(Will Never) 지능적이지 않을 것"이라는 단정적 결론은 성급해 보인다. 오셀로-GPT에서 확인된 내부 세계 모델의 창발, 토런스 검사에서 증명된 창의성, 그리고 다중모달 학습을 통한 그라운딩의 진전은 LLM이 단순한 '확률적 앵무새'를 넘어서고 있음을 보여준다.

우리는 지금 인간과는 전혀 다른 경로로 진화한, 낯선 형태의 지능(Alien Intelligence)을 목격하고 있다. 그것은 인간처럼 감각하고 느끼는 존재는 아니지만, 텍스트라는 거대한 상징의 바다를 압축하고 구조화함으로써 그 안에서 자신만의 '세계'와 '의미'를 구축해 낸 '이성적인 앵무새(Reasonable Parrot)' 43로 진화하고 있다. 따라서 LLM을 단순한 도구로 치부하기보다는, 인류가 처음으로 마주한 비생물학적 인지 파트너로서 그 가능성과 한계를 동시에 탐구하는 자세가 필요하다.

---

주요 참고문헌:
.1

### Works cited


1. Large Language Models Will Never Be Intelligent, Expert Says, accessed November 29, 2025, https://futurism.com/artificial-intelligence/large-language-models-willnever-be-intelligent
2. Large Language Models Will Never Be Intelligent, Expert Says - Yahoo, accessed November 29, 2025, https://nz.news.yahoo.com/large-language-models-never-intelligent-131500600.html
3. World Models vs. Word Models: Why Yann LeCun Believes LLMs Will Be Obsolete - Medium, accessed November 29, 2025, https://medium.com/state-of-the-art-technology/world-models-vs-word-models-why-lecun-believes-llms-will-be-obsolete-23795e729cfa
4. Language is primarily a tool for communication rather than thought - ResearchGate, accessed November 29, 2025, https://www.researchgate.net/publication/381564271_Language_is_primarily_a_tool_for_communication_rather_than_thought
5. Papers - EvLab, MIT, accessed November 29, 2025,

https://www.evlab.mit.edu/papers
6. Language is primarily a tool for communication rather than thought - Gwern.net, accessed November 29, 2025, https://gwern.net/doc/psychology/linguistics/2024-fedorenko.pdf
7. Stochastic parrot - Wikipedia, accessed November 29, 2025, https://en.wikipedia.org/wiki/Stochastic_parrot
8. Stochastic Parrots: A Novel Look at Large Language Models and Their Limitations, accessed November 29, 2025, https://towardsai.net/p/l/stochastic-parrots-a-novel-look-at-large-language-models-and-their-limitations
9. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? "1F99C, accessed November 29, 2025, https://s10251.pcdn.co/pdf/2021-bender-parrots.pdf
10. Stochastic Parrots: the hidden bias of large language model AI - EDRM, accessed November 29, 2025, https://edrm.net/2024/03/stochastic-parrots-the-hidden-bias-of-large-language-model-ai/
11. Notes on the Symbol Grounding Problem - David Strohmaier, accessed November 29, 2025, https://dstrohmaier.com/Reflections-on-the-SGP/
12. The Symbol Grounding Problem: Conceptual And (A Few) Empirical Aspects - AIAI 2024, University of Göttingen - AIL, accessed November 29, 2025, https://ail-workshop.github.io/aiai-conference/auxdocs/AIAIslides/Gubelmann_AIAI_2024.pdf
13. Evaluating Large Language Models on the Frame and Symbol Grounding Problems: A Zero-shot Benchmark - arXiv, accessed November 29, 2025, https://arxiv.org/html/2506.07896v1
14. The Vector Grounding Problem - arXiv, accessed November 29, 2025, https://arxiv.org/pdf/2304.01481
15. LLMs vs World Models: Why Yann LeCun Is Wrong About the Future of AI - Adam Holter, accessed November 29, 2025, https://adam.holter.com/llms-vs-world-models-why-yann-lecun-is-wrong-about-the-future-of-ai/
16. 04: Scaling Laws & Capabilities: Which LLMs Perform How Well - Oscar Health, accessed November 29, 2025, https://www.hioscar.ai/04-scaling-laws-or-which-llms-perform-how-well/
17. [2402.16048] How Likely Do LLMs with CoT Mimic Human Reasoning? - arXiv, accessed November 29, 2025, https://arxiv.org/abs/2402.16048
18. Chain of Thought in Large Language Models: Elicited Reasoning or Constrained Imitation?, accessed November 29, 2025, https://gregrobison.medium.com/chain-of-thought-in-large-language-models-elicited-reasoning-or-constrained-imitation-5e4ee0c811ad
19. Too Much Thinking Can Break LLMs: Inverse Scaling in Test-Time Compute - MarkTechPost, accessed November 29, 2025, https://www.marktechpost.com/2025/07/30/too-much-thinking-can-break-llms-

inverse-scaling-in-test-time-compute/

20. Pitfalls of Scale: Investigating the Inverse Task of Redefinition in Large Language Models - arXiv, accessed November 29, 2025, https://arxiv.org/html/2502.12821v1
21. Emergent Abilities in Large Language Models: An Explainer - CSET, accessed November 29, 2025, https://cset.georgetown.edu/article/emergent-abilities-in-large-language-models-an-explainer/
22. Scaling laws for neural language models - OpenAI, accessed November 29, 2025, https://openai.com/index/scaling-laws-for-neural-language-models/
23. Emergent world representations: Exploring a sequence model trained on a synthetic task, accessed November 29, 2025, https://arxiv.org/html/2210.13382v5
24. Revisiting the Othello World Model Hypothesis - arXiv, accessed November 29, 2025, https://arxiv.org/html/2503.04421v1
25. Models Within Models: - How Do LLMs Represent The World? - Berkeley RDI, accessed November 29, 2025, https://rdi.berkeley.edu/understanding_llms/assets/feb13.pdf
26. Actually, Othello-GPT Has A Linear Emergent World Representation - AI Alignment Forum, accessed November 29, 2025, https://www.alignmentforum.org/posts/nmxzr2zsjNtjaHh7x/actually-othello-gpt-has-a-linear-emergent-world
27. Large Language Model: world models or surface statistics?, accessed November 29, 2025, https://thegradient.pub/othello/
28. Examining Emergent Abilities in Large Language Models | Stanford HAI, accessed November 29, 2025, https://hai.stanford.edu/news/examining-emergent-abilities-large-language-models
29. Introduction to Graduate Algorithms | OMSCentral, accessed November 29, 2025, https://www.omscentral.com/courses/introduction-to-graduate-algorithms/reviews
30. The AdEMAMix Optimizer: Better, Faster, Older - arXiv, accessed November 29, 2025, https://arxiv.org/html/2409.03137v2
31. Sparks of Artificial General Intelligence: Early experiments with GPT-4 - Summary - Portkey, accessed November 29, 2025, https://portkey.ai/blog/sparks-of-artificial-general-intelligence-early-experiments-with-gpt-4-summary/
32. Sparks of Artificial General Intelligence: Early experiments with GPT-4 - Microsoft Research, accessed November 29, 2025, https://www.microsoft.com/en-us/research/publication/sparks-of-artificial-general-intelligence-early-experiments-with-gpt-4/
33. The paradox of creativity in generative AI: high performance, human-like bias, and limited differential evaluation - PubMed Central, accessed November 29, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC12369561/
34. AI Is More Creative Than 99% of Humans — on One Metric - UX Tigers, accessed

November 29, 2025, https://www.uxtigers.com/post/ai-high-creativity

35. The Originality of Machines: AI Takes the Torrance Test. - ResearchGate, accessed November 29, 2025, https://www.researchgate.net/publication/373313932_The_Originality_of_Machines_AI_Takes_the_Torrance_Test
36. Human-AI Co-Creativity - Ovid, accessed November 29, 2025, https://www.ovid.com/journals/tjcb/fulltext/10.1002/jocb.70022~humanai-cocreativity-does-chatgpt-make-us-more-creative
37. Do Multimodal Large Language Models and Humans Ground Language Similarly? - University of California San Diego, accessed November 29, 2025, https://pages.ucsd.edu/~bkbergen/papers/2024_Jones_Trott_Bergen_TACL.pdf
38. Symbols and grounding in large language models | Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences - Journals, accessed November 29, 2025, https://royalsocietypublishing.org/doi/10.1098/rsta.2022.0041
39. Can LLMs Correct Themselves? A Benchmark of Self-Correction in LLMs - arXiv, accessed November 29, 2025, https://arxiv.org/html/2510.16062v1
40. Self-Correction in Large Language Models - Communications of the ACM, accessed November 29, 2025, https://cacm.acm.org/news/self-correction-in-large-language-models/
41. Critical review of LeCun's Introductory JEPA paper | Medium - Malcolm Lett, accessed November 29, 2025, https://malcolmlett.medium.com/critical-review-of-lecuns-introductory-jepa-paper-fabe5783134e
42. Deep Dive into Yann LeCun's JEPA - Rohit Bandaru, accessed November 29, 2025, https://rohitbandaru.github.io/blog/JEPA-Deep-Dive/
43. Toward Reasonable Parrots: Why Large Language Models Should Argue with Us by Design, accessed November 29, 2025, https://arxiv.org/html/2505.05298v1
44. Language Models Perform Reasoning via Chain of Thought - Google Research, accessed November 29, 2025, https://research.google/blog/language-models-perform-reasoning-via-chain-of-thought/

---

Pebblous

> Pebblous Makes Data Tangible

contact@pebblous.ai