



# 차세대 AI를 위한 세 가지 월드 모델 비교: *Jeff Hawkins, Yann LeCun, Fei-Fei Li*

---

- 기획: 페블러스 데이터커뮤니케이션팀
- 감수: 이주행 (페블러스 대표)
- 작성일: 2026-01-02
- 공개여부: 공개

## 1. 서론: 언어를 넘어 세계로의 확장

---

인공지능(AI) 기술은 지난 10년간 심층 신경망(Deep Neural Networks), 특히 대규모 언어 모델(Large Language Models, LLMs)의 비약적인 발전을 통해 인류 지성사에 유례없는 변곡점을 맞이했다. OpenAI의 ChatGPT 시리즈로 대표되는 생성형 AI는 텍스트를 이해하고 생성하는 능력에서 인간에 근접하거나 능가하는 성능을 보여주었다. 그러나 이러한 언어 중심의 AI가 가진 근본적인 한계 또한 명확해지고 있다. **텍스트는 세계를 설명하는 상징(Symbol)일 뿐, 물리적 세계(Physical World) 그 자체는 아니다.** LLM은 방대한 텍스트 데이터의 통계적 상관관계를 학습하여 다음에 올 단어를 예측하는 데 탁월하지만, "왜" 그런 일이 일어나는지에 대한 인과적 추론(Causal Inference)이나, 자신의 행동이 물리적 환경에 어떤 결과를 초래할지에 대한 공간적 이해(Spatial Understanding)는 결여되어 있다. 이른바 "환각(Hallucination)" 현상은 AI가 현실 세계의 법칙에 기반하지 않고 확률적 그럴듯함에 의존하기 때문에 발생하는 필연적인 부작용이다 [1].

이러한 배경에서 AI 연구의 최전선은 이제 '언어 모델'에서 '**월드 모델(World Model)**'로 급격히 이동하고 있다. 월드 모델이란 에이전트(인간 또는 기계)가 외부 환경의 구조와 작동 원리를 내부적으로 시뮬레이션하여, 행동의 결과를 예측하고 계획을 수립할 수 있게 하는 인지적 프레임워크를 의미한다. 이는 단순히 데이터를 분류하거나

생성하는 것을 넘어, 지능의 본질인 '이해(Understanding)'와 '생존(Survival)'을 가능케 하는 핵심 기제다.

본 보고서는 월드 모델 연구를 주도하는 세 명의 세계적 석학, **Jeff Hawkins** (Numenta), **Yann LeCun** (Meta), **Fei-Fei Li** (World Labs)의 이론과 기술적 접근 방식을 심층적으로 분석한다. Jeff Hawkins는 뇌과학적 사실에 입각하여 인간의 신피질(Neocortex)이 세계를 모델링하는 메커니즘을 역공학(Reverse Engineering)한다. Yann LeCun은 현재의 생성형 AI가 가진 비효율성을 비판하며, 추상적 표현 공간(Representation Space)에서의 예측과 계획을 강조하는 인지 아키텍처를 제안한다. Fei-Fei Li는 시각적 인식을 넘어 3차원 공간을 구축하고 조작할 수 있는 '공간 지능(Spatial Intelligence)'을 통해 디지털과 물리적 세계의 융합을 시도한다.

이 세 가지 관점은 상충하는 듯 보이지만, 실상은 일반 인공지능(Artificial General Intelligence, AGI)이라는 거대한 목표를 향해 서로 다른 층위(Layer)에서 해답을 제시하고 있다. 본 보고서는 이들의 이론적 배경, 아키텍처의 기술적 세부 사항, 그리고 각 모델이 제시하는 미래 AI의 청사진을 면밀히 비교 분석함으로써, 차세대 AI가 나아가야 할 방향성을 제시하고자 한다.

## 2. Jeff Hawkins와 Numenta: 생물학적 실재론과 천 개의 뇌 이론

---

Jeff Hawkins는 팜(Palm) 컴퓨팅의 창업자로 유명하지만, 그의 진정한 열정은 지난 수십 년간 뇌의 작동 원리를 규명하고 이를 기반으로 진정한 지능을 구현하는 데 있었다. 그가 이끄는 **Numenta**는 신경과학적 발견을 수학적 알고리즘으로 변환하는 연구를 지속해왔으며, 이는 '**천 개의 뇌 이론(Thousand Brains Theory)**'으로 집대성되었다.

### 2.1 천 개의 뇌 이론 (The Thousand Brains Theory)의 핵심

기존의 신경과학과 AI 이론은 뇌가 감각 입력을 계층적으로 처리하여 최상위 계층에서 단일한 객체 모델을 형성한다고 가정했다. 즉, 시각 정보가 V1, V2, V4 등을 거쳐 IT(Inferior Temporal) 피질에 도달해서야 '컵'이라는 객체를 인식한다는 것이다. 그러나 Hawkins는 이러한 계층적 통합설을 부정하고, **분산된 모델링(Distributed Modeling)** 가설을 제시한다 [3].

## 2.1.1 피질 기둥 (Cortical Column)의 독립성

Hawkins의 이론에 따르면, 신피질을 구성하는 약 15만 개의 **피질 기둥(Cortical Columns)** 각각은 완전한 감각운동 모델링 시스템이다. 뇌의 특정 영역만이 모델링을 담당하는 것이 아니라, 모든 피질 기둥이 각자 독립적으로 입력된 감각 정보를 바탕으로 세계에 대한 모델을 학습한다. 예를 들어, 커피잔을 잡고 있는 다섯 손가락은 각각 서로 다른 감각 입력을 받지만, 각 손가락에 연결된 피질 기둥들은 독자적으로 "이것은 커피잔이다"라는 모델을 형성하려고 시도한다 [5].

## 2.1.2 투표 메커니즘 (Voting Mechanism)과 바인딩 문제의 해결

그렇다면 수만 개의 독립된 모델이 어떻게 하나의 통일된 지각(Perception)을 형성하는가? Hawkins는 이를 '**투표(Voting)**'로 설명한다. 피질 기둥들은 신피질의 장거리 신경 연결(Long-range connections)을 통해 서로 정보를 교환한다.

- **상황:** 검지 손가락은 '매끄러운 곡면'을 느끼고, 엄지 손가락은 '손잡이'를 느낀다.
- **개별 판단:** 검지 기둥은 이것이 컵, 공, 또는 사과일 수 있다고 판단한다. 엄지 기둥은 이것이 컵 또는 주전자일 수 있다고 판단한다.
- **합의(Consensus):** 두 기둥이 정보를 교환하면, 유일한 공통 분모인 '컵'으로 의견이 모아진다. 이 순간 우리는 "이것은 컵이다"라고 인식하게 된다 [6].

이러한 방식은 뇌가 손상되거나 노이즈가 많은 환경에서도 강건하게 작동할 수 있는 이유를 설명한다. 일부 기둥이 손상되어도 나머지 기둥들의 투표를 통해 정확한 인식이 가능하기 때문이다.

## 2.2 참조 프레임 (Reference Frames): 지능의 좌표계

Hawkins 이론의 가장 혁신적인 부분은 뇌가 정보를 저장하고 처리하는 기본 형식이 '**참조 프레임(Reference Frame)**'이라는 발견이다. 그는 뇌의 '격자 세포(Grid Cells)'와 '장소 세포(Place Cells)'가 해마뿐만 아니라 신피질 전체에 존재한다고 주장한다 [8].

### 2.2.1 객체 중심 좌표계 (Allocentric Representation)

우리가 방 안에서 움직여도 방의 구조가 바뀌지 않는 것처럼, 뇌는 관찰자(자아) 중심이

아닌 **객체 중심(Allocentric)** 의 좌표계를 사용하여 대상을 모델링한다.

- **What (특징):** 감각 기관이 느끼는 특징(Feature).
- **Where (위치):** 그 특징이 객체의 참조 프레임 내에서 어디에 위치하는지 (Relative Location).

피질 기둥의 L4(Layer 4) 층은 감각 입력을 받고, L6(Layer 6) 층은 현재 센서의 위치 정보를 처리한다. 이 두 정보가 결합하여 "이 객체의 이 위치에 이러한 특징이 있다"라는 지식이 형성된다 [8]. 이는 기존의 딥러닝이 픽셀 패턴(Texture) 위주로 학습하는 것과 달리, 뇌는 철저히 **구조(Structure)**와 **기하학(Geometry)** 을 학습한다는 것을 시사한다.

### 2.2.2 추상적 사고로의 확장

Hawkins는 이러한 참조 프레임 메커니즘이 물리적 객체뿐만 아니라 수학, 언어, 정치와 같은 **추상적 개념(Abstract Concepts)** 을 이해하는 데에도 동일하게 적용된다고 주장한다. 민주주의라는 개념도 뇌 속에서는 관련된 하위 개념들이 특정한 논리적 '위치'와 '관계'를 맺고 있는 구조물로 저장된다는 것이다 [3].

## 2.3 감각운동 통합 (Sensorimotor Integration)과 학습

Hawkins는 "**생각하는 행위 자체가 움직임의 한 형태**"라고 말한다. 우리는 가만히 대상을 바라보는 것이 아니라, 눈을 움직이고(Saccade), 손으로 만지며 정보를 수집한다 [4].

- **예측(Prediction):** 뇌는 끊임없이 다음 감각 입력을 예측한다. "내가 손가락을 오른쪽으로 움직이면, 컵의 손잡이가 느껴질 것이다"라는 예측이 맞으면 모델이 강화되고, 틀리면 수정된다(Predictive Coding) [6].
- **행동(Action):** 따라서 행동은 학습과 분리될 수 없다. Hawkins의 모델에서 지능은 본질적으로 **\*\*구체화(Embodied)\*\***되어 있으며, 감각과 운동 시스템은 하나의 루프(Loop)로 통합되어 있다 [5].

## 2.4 희소 분산 표현 (Sparse Distributed Representations, SDR)

Numenta 기술의 기저에는 **SDR**이라는 데이터 구조가 있다. 이는 뇌의 뉴런이 매우 희소하게(Sparsely) 활성화되는 것을 모방한 것이다. 2,048개의 비트 중 약 2%(40개)만이 1이 되고 나머지는 0이 되는 식이다 [11].

- **의미론적 내포:** SDR의 각 비트는 의미를 가진다. 두 SDR이 겹치는 비트가 많을수록 의미적으로 유사하다.
- **강건성:** 40개의 활성 비트 중 절반이 노이즈로 인해 사라져도, 나머지 20개를 통해 전체 패턴을 유추할 수 있다. 이는 생물학적 뇌의 놀라운 노이즈 내성을 설명한다 [12].

### 3. Yann LeCun과 Meta: 자율 기계 지능을 위한 인지 아키텍처

---

튜링상 수상자이자 Meta의 수석 AI 과학자인 Yann LeCun은 현재 AI 업계를 지배하는 LLM의 자기회귀(Auto-regressive) 방식에 대해 가장 강력한 비판자 중 한 명이다. 그는 진정한 AGI로 나아가기 위해서는 텍스트 생성 모델이 아닌, 추론하고 계획할 수 있는 '**월드 모델**'이 필요하다고 역설하며 **JEPA(Joint Embedding Predictive Architecture)**를 제안했다.

#### 3.1 생성형 AI에 대한 비판: "LLM은 월드 모델이 아니다"

LeCun은 텍스트나 픽셀을 순차적으로 예측하는 방식이 근본적인 한계에 봉착했다고 주장한다 [1].

- **비효율성:** 비디오 생성 모델(예: Sora)이 3D 세계를 시뮬레이션하는 것처럼 보이지만, 이는 픽셀 공간에서의 통계적 모방일 뿐이다. LeCun은 "월드 모델을 만들기 위해 픽셀을 생성하는 것은 마치 물리학을 배우기 위해 양자 역학 수준의 모든 입자 위치를 계산하는 것과 같이 비효율적"이라고 비판한다 [15].
- **오류 누적:** 자기회귀 모델은 생성된 결과물을 다시 입력으로 사용하기 때문에, 초기 오류가 기하급수적으로 증폭되어 현실과 동떨어진 환각(Drift)을 만들어낸다 [16].
- **데이터 대역폭의 차이:** 인간은 텍스트(초당 몇 바이트)보다 시각 정보(초당 수십 메가바이트)를 통해 훨씬 더 많은 정보를 받아들이며 세계를 배운다. 텍스트만으로

학습한 AI는 물리적 상식(Common Sense)을 가질 수 없다[17].

## 3.2 JEPA (Joint Embedding Predictive Architecture): 추상적 공간에서의 예측

LeCun의 해법은 입력 공간(Input Space)이 아닌 표현 공간(Representation Space)에서 예측을 수행하는 것이다[18].

### 3.2.1 JEPA의 작동 원리

JEPA는 입력  $x$ (현재 상태)와  $y$ (미래 상태)를 직접 연결하지 않는다.

1. **인코더(Encoder):** 입력  $x$ 와  $y$ 를 각각 추상적인 표현 벡터  $s_x, s_y$ 로 변환한다. 이때 인코더는 예측에 불필요한 세부 정보(예: 배경의 흔들리는 나뭇잎, 카메라 노이즈)를 제거하고 중요한 정보(예: 물체의 위치, 종류, 운동량)만을 남긴다 [19].
2. **예측기(Predictor):**  $s_x$ 와 잠재 변수  $z$ 를 바탕으로  $s'_y$ 를 예측한다. 목표는 예측된  $s'_y$ 가 실제  $s_y$ 와 최대한 가까워지도록 하는 것이다.
3. **비생성적(Non-generative):** 중요한 점은 JEPA가  $s'_y$ 를 다시 픽셀 이미지로 복원(Decoding)하지 않는다는 것이다. 예측은 오직 추상 공간에서만 이루어진다.

### 3.2.2 에너지 기반 모델 (Energy-Based Models, EBM)

LeCun의 이론적 토대는 EBM이다. 학습은 두 상태(입력과 예측) 사이의 '에너지(불일치도/비용)'를 최소화하는 방향으로 이루어진다.

- **호환성(Compatibility):** 관찰된 데이터  $x$ 와  $y$ 가 물리적으로 호환되면 낮은 에너지를, 불가능한 상황이면 높은 에너지를 부여한다 [20].
- **불확실성 관리:** 미래는 결정되어 있지 않다(Stochastic). JEPA는 잠재 변수  $z$ 를 사용하여 여러 가능한 미래를 모델링한다. 생성 모델처럼 모든 픽셀의 평균값을 예측하여 흐릿한(Blurry) 이미지를 만드는 대신,  $z$ 값에 따라 타당한 여러 미래 상태를 명확한 벡터로 예측할 수 있다 [18].

## 3.3 계층적 계획 (Hierarchical Planning)과 자율 에이전트

LeCun이 구상하는 자율 지능 시스템은 6개의 모듈로 구성된다 [18].

1. **구성기(Configurator)**: 작업 목표를 설정한다.
2. **인식(Perception)**: 센서 입력을 받아 현재 상태를 추정한다.
3. **월드 모델(World Model)**: 행동에 따른 미래 상태를 예측한다.
4. **비용(Cost)**: 본능적인 비용(Intrinsic Cost, 예: 충돌 회피)과 학습된 비평가(Critic)를 통해 상태의 가치를 평가한다.
5. **단기 기억(Short-term Memory)**: 현재 상황과 계획을 저장한다.
6. **행동기(Actor)**: 비용을 최소화하는 행동 시퀀스를 생성한다.

여기서 **계층적 JEPA (H-JEPA)** 는 추상화 수준을 달리하여 장기 계획을 가능하게 한다. 상위 레벨에서는 "공항으로 간다"는 거시적 계획을, 하위 레벨에서는 "핸들을 왼쪽으로 10도 꺾는다"는 미시적 제어를 처리한다 [18]. 이는 인간의 **시스템 2(System 2)** 사고, 즉 느리고 신중한 계획 능력을 AI에 구현하려는 시도이다 [23].

### 3.4 I-JEPA와 V-JEPA

Meta는 이 이론을 입증하기 위해 이미지용 **I-JEPA**와 비디오용 **V-JEPA**를 공개했다. I-JEPA는 이미지의 일부를 가리고(Masking) 주변 정보를 통해 그 부분의 '의미적 표현'을 예측하며, V-JEPA는 비디오의 시간적 흐름 속에서 보이지 않는 부분의 시공간적 특징을 예측한다. 이들은 픽셀 생성 없이도 객체 인식과 물리적 상호작용 이해에서 뛰어난 성능(Sample Efficiency)을 보여주었다 [20].

## 4. Fei-Fei Li와 World Labs: 공간 지능과 3D 생성의 융합

---

Fei-Fei Li는 ImageNet을 통해 컴퓨터 비전의 혁명을 촉발한 인물이다. 그녀는 이제 "보는 것(Seeing)"을 넘어 "행동하는 것(Doing)"으로 AI의 패러다임을 확장하기 위해 World Labs를 설립하고 '**공간 지능(Spatial Intelligence)**'이라는 개념을 주창했다.

### 4.1 공간 지능 (Spatial Intelligence): AI의 잃어버린 조각

Li는 현재의 AI가 언어와 2D 이미지 생성에는 능통하지만, 3차원 공간을 이해하고 조작



하는 능력은 현저히 부족하다고 지적한다. 그녀는 이를 "어둠 속의 언어술사 (Wordsmiths in the dark)"라고 표현하며, 현실 감각이 결여된 지능의 한계를 꼬집었다.<sup>24</sup>

- **정의:** 공간 지능은 "상상(Imagination), 지각(Perception), 행동(Action)"을 연결하는 능력이다. 이는 3D 공간의 깊이, 거리, 물리적 상호작용, 시간적 변화를 이해하는 것을 포함한다 [26].
- **진화적 관점:** 생명체의 지능은 시각적 인식을 통해 환경과 상호작용하며 진화했다. 캄브리아기 대폭발이 '눈'의 탄생으로 시작되었듯, AI의 다음 진화는 3D 공간을 이해하는 능력에서 비롯될 것이다 [24].

## 4.2 Marble: 대형 월드 모델 (Large World Model, LWM)

World Labs의 첫 번째 결과물인 **Marble**은 텍스트, 이미지, 비디오 등 다양한 입력을 받아 완전한 3D 상호작용 환경을 생성하는 생성형 월드 모델이다 [27].

### 4.2.1 명시적 3D 표현 (Explicit 3D Representation)

LeCun의 JEPA가 내부적인 추상 표현을 지향하는 것과 달리, Marble은 사용자가 탐색하고 상호작용할 수 있는 **명시적인 3D 자산**을 생성한다.

- **Gaussian Splats (가우시안 스플랫):** Marble은 3D 장면을 수많은 반투명 입자 (Gaussian)의 집합으로 표현하여, 사진과 같은 고해상도(Photorealistic) 시각화를 실시간으로 렌더링한다 [27].
- **Meshes & Colliders:** 시각적 표현뿐만 아니라 물리 엔진이 인식할 수 있는 충돌 메쉬(Collider Mesh)를 함께 생성한다. 이는 생성된 세계가 단순한 배경 그림이 아니라, 로봇이나 게임 캐릭터가 걸어 다니고 물체와 부딪힐 수 있는 물리적 공간임을 의미한다 [29].

### 4.2.2 생성 기술의 혁신: 구조와 스타일의 분리

Marble은 생성 AI의 고질적인 문제인 '제어 불가능성(Uncontrollability)'을 해결하기 위해 구조와 스타일을 분리하는 접근을 취한다.

- **Chisel (조각 모드):** 사용자가 기본적인 3D 구조(박스, 평면 등)를 배치하면, AI가



그 위에 텍스트 프롬프트에 맞는 텍스처와 디테일을 입힌다 [27].

- **멀티모달 입력 통합:** 여러 장의 사진이나 비디오를 입력하면 이를 정합(Stitching)하여 일관된 하나의 3D 공간으로 재구성한다. 이는 부동산, 건축, 게임 레벨 디자인 등에서 즉각적인 유용성을 가진다 [31].

### 4.3 RTFM (Real-Time Frame Model): 학습된 렌더러

최근 공개된 **RTFM**은 Marble과는 또 다른 접근을 보여준다. 이는 명시적인 3D 지오메트리(Mesh 등)를 생성하지 않고, **신경망 자체가 3D 세계를 암묵적으로 기억하고 렌더링하는 방식**이다 [32].

- **작동 방식:** RTFM은 비디오 데이터를 학습하여, 입력된 이미지의 새로운 시점(Viewpoint)을 예측한다. 이는 트랜스포머 기반의 자기회귀 모델이지만, 3D 유클리드 공간의 사전 지식(Prior)을 가지고 있어 '공간적 일관성'을 유지한다.
- **컨텍스트 저글링(Context Juggling):** 대규모 세계를 탐색할 때 메모리 폭발을 막기 위해, 현재 시점과 관련된 프레임만을 동적으로 호출하여 지속성(Persistence)을 유지한다. 이는 LeCun의 '추상적 시뮬레이션'과 Li의 '시각적 생성' 사이의 가교 역할을 하는 기술로 평가받는다.

### 4.4 로보틱스와 Sim2Real의 혁명

Li의 비전은 궁극적으로 로보틱스로 향한다. 현실 세계에서 로봇을 학습시키는 것은 느리고 위험하다. Marble로 생성된 무한하고 다양한 3D 시뮬레이션 환경(Sim)에서 로봇을 학습시키고, 이를 현실(Real)로 가져오는 **Sim2Real** 전략은 로봇 공학의 데이터 병목 현상을 해결할 열쇠로 꼽힌다 [29].

## 5. 비교 분석: 세 가지 이론의 통합과 대립

세 석학의 이론은 모두 '현재의 AI 한계 극복'과 '물리적 세계의 이해'를 목표로 하지만, 그 접근 방식(Approach)과 존재론적 관점(Ontology)에서 뚜렷한 차이를 보인다.

### 5.1 비교 요약표

--	--	--	--

비교 항목	Jeff Hawkins ≡(Numenta)	Yann LeCun ≡(AMI Lab)	Fei-Fei Li ≡(World Labs)
핵심 이론	천 개의 뇌 이론 (Thousand Brains)	자율 기계 지능 ≡(JEPA)	공간 지능 ≡(Spatial Intelligence)
주요 아키텍처	참조 프레임 & 피질 기둥 (SDR)	Joint Embedding & 예측기	생성적 3D 모델 (Marble, RTFM)
월드 모델의 정의	피질 기둥에 분산된 수천 개의 센서-위치 모델	추상적 표현 공간에서의 상태 예측 시뮬레이터	물리적 상호작용이 가능한 명시적 3D 환경
데이터 표현	희소 분산 표현 (SDR), Grid Cells	고차원 임베딩 벡터 (Latent Representations)	Gaussian Splats, Meshes, Neural Rendering
생성 (Generation)	내부적 예측 (Predictive Coding)을 위한 생성	픽셀 생성 반대 (비효율적, 환각 위험)	적극적 생성 (Generative AI)을 통한 공간 구축
물리학의 구현	감각운동 학습을 통해 구조적으로 체득	에너지 함수 최소화를 통해 직관적으로 학습	물리 엔진과의 결합 또는 3D 시뮬레이션 제공
지향점 (Goal)	생물학적 뇌의 역공학을 통한 AGI	인간 수준의 추론/계획 능력을 가진 AI	디지털-물리 융합 및 로봇틱스 생태계

### 5.2 논쟁점 1: 표현(Representation) - 구체성 vs. 추상성

- **Hawkins vs. LeCun:** Hawkins는 뇌가 '지도(Map)'와 같은 기하학적 구조(참조 프레임)를 실제로 가지고 있다고 본다. 정보는 '위치'와 결합되어 저장된다. 반면 LeCun은 정보가 수학적으로 효율적인 '추상 공간'에 압축되어야 한다고 본다. 그러나 LeCun의 JEPA에서 잠재 변수( $z$ )가 공간적 불확실성을 처리하는 방식은 Hawkins의 투표 메커니즘이 불확실성을 해소하는 방식과 기능적으로 유사하다.
- **Li vs. LeCun:** LeCun은 픽셀이나 3D 좌표를 직접 생성하는 것을 "낭비"라고 본다. 지능은 컵의 정확한 3D 메쉬를 그리는 것이 아니라, 컵을 잡을 수 있는 '개념적 이해'만 있으면 된다는 것이다 [15]. 반면 Li는 로봇이 실제로 행동하기 위해서는 "구체적인 3D 환경"이 필수적이라고 본다. Li의 Marble은 LeCun이 거부한 '세밀한 픽셀/복셀의 생성'을 통해 역설적으로 로봇에게 '안전한 훈련장'을 제공한다.

### 5.3 논쟁점 2: 생성(Generation)의 역할

- **LeCun의 비판:** Sora와 같은 비디오 생성 모델은 물리 법칙을 이해한 것이 아니라 흉내 내는 것이다. 따라서 엔지니어링 도구로는 훌륭할지 몰라도, 지능 모델로서는 실패할 운명이다.
- **Li의 반박적 수용:** Li의 모델은 생성형이지만, 2D 픽셀의 통계적 나열이 아니라 3D 기하학적 일관성(Geometric Consistency)을 강제한다. 즉, Marble은 '보기에 그럴듯한' 영상이 아니라 '물리적으로 타당한' 3D 구조를 생성함으로써 LeCun의 비판(물리적 이해 부재)을 기술적으로 극복하려 한다 [31].
- **Hawkins의 관점:** 뇌는 끊임없이 입력을 생성(예측)한다. 그러나 이는 외부 출력을 위한 것이 아니라 내부 모델의 검증을 위한 것이다. Hawkins에게 생성은 학습의 수단이지 목적이 아니다.

### 5.4 상호보완성: AGI를 향한 융합

이 세 이론은 상호 배타적이라기보다 보완적이다.

1. **구조(Structure):** Hawkins의 **참조 프레임** 이론은 AI가 정보를 저장하는 '데이터 구조'에 대한 청사진을 제공한다. 트랜스포머의 어텐션 메커니즘도 일종의 암묵적 참조 프레임으로 볼 수 있다.
2. **추론(Reasoning):** LeCun의 **JEPA**는 이 구조 위에서 작동하는 '추론 엔진'이다. 불필요한 정보를 버리고 핵심 인과관계만을 학습하여 효율적으로 계획을 수립한다.

3. **환경(Environment):** Li의 **Marble**은 이 AI가 학습하고 상호작용할 '신체와 환경'을 제공한다. JEPA를 탑재한 로봇(Brain)이 Marble로 생성된 세계(World)에서 Hawkins의 방식(Sensorimotor)으로 학습하는 미래를 그려볼 수 있다.

## 6. 결론: 월드 모델이 여는 미래

---

Jeff Hawkins, Yann LeCun, Fei-Fei Li는 각기 다른 출발점에서 시작했지만, 결론은 하나의 지점으로 수렴하고 있다. "지능은 정적인 데이터를 학습하는 것이 아니라, 역동적인 세계의 구조와 인과관계를 내재화하는 것이다."

- **Jeff Hawkins**는 우리에게 뇌가 세계를 어떻게 **\*\*좌표화\*\***하는지 알려주었다. 그의 이론은 AI가 단순한 패턴 매칭을 넘어, 인간처럼 개념을 구조적으로 조작할 수 있는 길을 열어준다.
- **Yann LeCun**은 AI가 어떻게 **'효율적으로'** 생각해야 하는지 제시했다. 그의 비생성적 예측 모델은 AI가 불필요한 계산의 늪에 빠지지 않고, 인간 수준의 상식과 계획 능력을 갖추도록 유도한다.
- **Fei-Fei Li**는 AI에게 **\*\*살아갈 공간\*\***을 만들어주었다. 그녀의 공간 지능 기술은 디지털 지능을 물리적 현실로 확장시키며, 로봇틱스와 메타버스의 경계를 허물고 있다.

지금 우리는 텍스트 기반의 '챗봇' 시대에서, 물리적 현실을 이해하고 행동하는 '에이전트' 시대로 넘어가는 과도기에 있다. 이 세 석학이 구축하고 있는 월드 모델들은 그 새로운 시대를 지탱할 이론적, 기술적 기둥이 될 것이다. 향후 AI의 발전은 이 세 가지 흐름이 어떻게 융합되어, **신체(Embodiment)**를 가진, **추론(Reasoning)**하는, **구조적(Structured)** 지능으로 탄생하느냐에 달려 있다.

## 참고문헌

---

1. OpenAI's Video-Generating AI Is "Doomed to Failure," Says Meta's Top AI Scientist, accessed December 29, 2025, <https://futurism.com/the-byte/openai-video-ai-doomed-meta-scientist>
2. LLMs Were Just the Warm-Up. AI's Next Revolution is World Models - Yan Nuriyev, accessed December 29, 2025,

<https://whoisyan.com/llms-were-just-the-warm-up-ais-next-revolution-is-world-models/>

3. A Thousand Brains: A New Theory Of Intelligence by Jeff Hawkins - Numenta, accessed December 29, 2025, <https://www.numenta.com/resources/books/a-thousand-brains-by-jeff-hawkins/>
4. A Thousand Brains - Internet Archive, accessed December 29, 2025, [https://dn790007.ca.archive.org/0/items/artificial-intelligence/2021\\_a\\_thousand\\_brains-a\\_new\\_theory\\_of\\_intelligence-Jeff%20Hawkins%20%282021%29.pdf](https://dn790007.ca.archive.org/0/items/artificial-intelligence/2021_a_thousand_brains-a_new_theory_of_intelligence-Jeff%20Hawkins%20%282021%29.pdf)
5. A Thousand Brains Theory: A Review | by Christophe Pere | TDS Archive - Medium, accessed December 29, 2025, <https://medium.com/data-science/a-thousand-brains-theory-a-review-3ea6bbeeced0>
6. The Thousand Brains Theory of Intelligence - Numenta, accessed December 29, 2025, <https://www.numenta.com/blog/2019/01/16/the-thousand-brains-theory-of-intelligence/>
7. "A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex" - Numenta, accessed December 29, 2025, <https://www.numenta.com/assets/pdf/research-publications/papers/Companion-paper-to-Thousand-Brains-Theory-of-Intelligence.pdf>
8. Untitled - Microsoft, accessed December 29, 2025, [https://www.microsoft.com/en-us/research/wp-content/uploads/2019/03/42804\\_The\\_Thousand\\_Brains\\_Theory.pdf](https://www.microsoft.com/en-us/research/wp-content/uploads/2019/03/42804_The_Thousand_Brains_Theory.pdf)
9. The Thousand Brains Project - Numenta, accessed December 29, 2025, [https://www.numenta.com/wp-content/uploads/2024/06/Short\\_TBP\\_Overview.pdf](https://www.numenta.com/wp-content/uploads/2024/06/Short_TBP_Overview.pdf)
10. Numenta Publishes a New Theory on Sensorimotor Inference, accessed December 29, 2025, <https://www.numenta.com/press/2017/11/15/numenta-publishes-new-sensorimotor-theory-in-frontiers/>

11. Sparse Distributed Representations - Numenta, accessed December 29, 2025, <https://www.numenta.com/assets/pdf/biological-and-machine-intelligence/BaMI-SDR.pdf>
12. The HTM Spatial Pooler—A Neocortical Algorithm for Online Sparse Distributed Coding, accessed December 29, 2025, <https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2017.00111/full>
13. The HTM Spatial Pooler – a neocortical algorithm for online sparse distributed coding - bioRxiv, accessed December 29, 2025, <https://www.biorxiv.org/content/10.1101/085035v1.full.pdf>
14. World Models vs. Word Models: Why Yann LeCun Believes LLMs Will Be Obsolete, accessed December 29, 2025, <https://medium.com/state-of-the-art-technology/world-models-vs-word-models-why-lecun-believes-llms-will-be-obsolete-23795e729cfa>
15. Yann LeCun doubles down, claims Sora doesn't count : r/singularity - Reddit, accessed December 29, 2025, [https://www.reddit.com/r/singularity/comments/1atk19r/yann\\_lecun\\_doubles\\_down\\_claims\\_sora\\_doesnt\\_count/](https://www.reddit.com/r/singularity/comments/1atk19r/yann_lecun_doubles_down_claims_sora_doesnt_count/)
16. Yann LeCun: Intense Complaints Before Leaving the Company - 36氪, accessed December 29, 2025, <https://eu.36kr.com/en/p/3605931513267201>
17. Highlights from Lex Fridman's interview of Yann LeCun - LessWrong, accessed December 29, 2025, <https://www.lesswrong.com/posts/bce63kvsAMcwxPipX/highlights-from-lex-fridman-s-interview-of-yann-lecun>
18. A Path Towards Autonomous Machine Intelligence - Temple CIS, accessed December 29, 2025, <https://cis.temple.edu/tagit/presentations/A%20Path%20Towards%20Autonomous%20Machine%20Intelligence.pdf>
19. What is Joint Embedding Predictive Architecture (JEPA)? - Turing Post, accessed December 29, 2025, <https://www.turingpost.com/p/jepa>

20. Deep Dive into Yann LeCun's JEPA - Rohit Bandaru, accessed December 29, 2025, <https://rohitbandaru.github.io/blog/JEPA-Deep-Dive/>
21. LeCun's 2022 paper on autonomous machine intelligence rehashes but does not cite essential work of 1990-2015, accessed December 29, 2025, <https://people.idsia.ch/~juergen/lecun-rehash-1990-2022.html>
22. JEPA: LeCun's Path Towards More Human-Like AI | by Anil Jain - Medium, accessed December 29, 2025, <https://medium.com/@anil.jain.baba/jepa-lecuns-path-towards-more-human-like-ai-9535e48b3c65>
23. Critical review of LeCun's Introductory JEPA paper | Medium - Malcolm Lett, accessed December 29, 2025, <https://malcolmllett.medium.com/critical-review-of-lecuns-introductory-jepa-paper-fabe5783134e>
24. Viral! Fei-Fei Li's 10,000-Word Article Defines Next Decade of AI - 36氪, accessed December 29, 2025, <https://eu.36kr.com/en/p/3548078081093508>
25. Are World Models the Future of AI? - Blockchain Council, accessed December 29, 2025, <https://www.blockchain-council.org/ai/world-models-future-of-ai/>
26. Building Spatial Intelligence: How World Labs is Creating the Next ..., accessed December 29, 2025, <https://radical.vc/building-spatial-intelligence-how-world-labs-is-creating-the-next-frontier-in-ai/>
27. Marble: A Multimodal World Model - World Labs, accessed December 29, 2025, <https://www.worldlabs.ai/blog/marble-world-model>
28. New Marble AI Creates Entire 3D Worlds from Simple Text Prompts - Analytics Vidhya, accessed December 29, 2025, <https://www.analyticsvidhya.com/blog/2025/11/marble-world-ai-creates-3d-worlds-from-text/>
29. Scaling Robotic Simulation with Marble | World Labs, accessed December 29, 2025, <https://www.worldlabs.ai/case-studies/1->



## robotics

30. Simulate Robotic Environments Faster with NVIDIA Isaac Sim and World Labs Marble, accessed December 29, 2025, <https://developer.nvidia.com/blog/simulate-robotic-environments-faster-with-nvidia-isaac-sim-and-world-labs-marble/>
  31. World Models — Not Next, Current Frontier | by evoailabs | Nov, 2025 | Medium, accessed December 29, 2025, <https://evoailabs.medium.com/world-models-not-next-current-frontier-547c5eeb1307>
  32. RTFM: A Real-Time Frame Model | World Labs, accessed December 29, 2025, <https://www.worldlabs.ai/blog/rtfm>
- 

# Pebblous

**Pebblous Makes Data Tangible**

[contact@pebblous.ai](mailto:contact@pebblous.ai)