

Beyond LLMs: The Rise of Human Brain Clones and Digital Brain Competitors

VERSION: 1.0 (FIRST EDITION)

ASSEM HIJAZI

Contents

Beyond LLMs: The Rise of Human Brain Clones and Digital Brain Competitors	3
Executive Summary	4
Beyond LLMs: A New Wave of AI Systems.....	7
Categorizing the Alternatives: Grouping the New AI Systems.....	10
1. QRI – Alpha Centauri Group	12
2. JEPA – Meta AI	14
3. IBM NorthPole – IBM Research	16
4. MIT Liquid Neural Networks – MIT CSAIL.....	18
5. Spaun – Stanford (via Nengo / Chris Eliasmith's Lab)	20
6. Numenta HTM – Numenta (Jeff Hawkins).....	22
7. OpenCog Hyperon – SingularityNET / Ben Goertzel	24
8. DeepMind Gato – DeepMind (Google).....	26
9. Intel Loihi + INRC – Intel Neuromorphic Research Community	28
10. AIGO – Peter Voss / Adaptive AI Inc.	30
11. Mamba – CMU & Princeton	32
12. Cicero – Meta AI	34
13. LAMs – Arion Research (Language-Action Models)	36
14. LCMs – Meta AI (Language-Computation Models)	38
15. MRKL Systems – AI21 Labs	40
16. CoALA – Princeton / Stanford (Cognitive Architecture for Language and Action).....	42
17. Cognitive AI – Nova Spivack (The "Mind OS" Concept)	44
18. LLM-ACTR – Cognitive Science + LLM Hybrid Model	46
19. OGI Framework – Open General Intelligence.....	48
20. Nengo – Applied Brain Research.....	50
21. GFlowNet – DeepMind / Bengio Lab	52
22. vLLM – Virtual Large Language Model	54
23. Perceiver – DeepMind (Google)	56
24. Liquid-1 – Liquid AI	58
25. JEIR – Jina Embodied Intelligence Reasoner (by Jina AI)	60
26. Akida – BrainChip Holdings Ltd.	62
27. MENTA – NEUROTECH.ai	64
28. Recursive Cortical Network (RCN) – Vicarious (Acquired by Elon Musk’s Tesla AI team)	66

Beyond LLMs: The Rise of Human Brain Clones and Digital Brain Competitors

[Assem Hijazi – Experience, Knowledge & Ideas](#)

We are living in an era where many believe that AI has already reached its peak—or at least the midpoint of its journey. But in reality, we're still just getting started.

The true ambition behind artificial intelligence has always been to **replicate the human brain**—its ability to think, remember, create, organize, and adapt. Among the most well-known approaches today are **Large Language Models (LLMs)**, which gained massive attention largely due to the success of ChatGPT. But let's be clear: LLMs are not the only path toward brain-like intelligence—and certainly not the final form of it.

The AI ecosystem plays a huge role in shaping public perception. The dominance of LLMs today is driven largely by industry giants like **OpenAI, Microsoft, Google**, and others adopting and scaling these models. That doesn't mean LLMs are the best option—or even the most aligned with human cognition.

In fact, many alternative models and architectures are emerging, some from **leading universities like MIT and Stanford**, and others from **pioneering companies like IBM, Intel, DeepMind**, and smaller AGI startups. These systems aim to replicate different aspects of human intelligence, using techniques far beyond language prediction—like spiking neural networks, symbolic reasoning, cognitive planning, and even causal logic trees.

In this article, I've compiled a list of **28 distinct AI systems**, each designed to simulate the human brain in its own way. I haven't tested them all, and some remain in early research or conceptual stages—but they show clearly that **LLMs are not alone**, and more importantly, **not unbeatable**. The AI race is just beginning, and the future may surprise us all.

Executive Summary

In this summary, I'll explain why we listed these 28 models, what's happening in the AI field beyond LLMs, and where the future might lead.

Let's start with something essential: **the future of AI depends on three pillars—hardware, software, and the ecosystem that connects them.** And if I had to summarize it in one sentence: **the real winner will be whoever builds the AI chips that best clone the human brain.**

Hardware + AI Chips

Some models in this list are already deeply tied to **neuromorphic hardware**—chips designed to work like the brain. For example:

- **IBM NorthPole** and **Intel Loihi** focus on **spiking neural networks** that fire like real neurons.
- **Akida by BrainChip** is an **edge-ready chip** designed for ultra-low power AI in real-world devices like wearables and cameras.
- **Nengo** can run massive **brain simulations** on neuromorphic platforms—true digital brain experiments.

These are not chatbots. They are **hardware-powered intelligence engines**, and they're showing us what AI might look like when it's embedded into the world around us.

Domain-Specific vs General AI

We must also understand that **AI is inherently domain-specific**. Not all models are built to do everything. Some are designed for:

- **Strategic interaction** (like **Cicero** from Meta, trained on the Diplomacy game),
- **Robotics and motion** (like **MIT Liquid Neural Nets** or **DeepMind Gato**),
- **Symbolic reasoning** in decision tools (like **OpenCog Hyperon** or **MRKL by AI21 Labs**),
- Or even **smart cities and edge AI** using hardware like **Akida** and **Loihi**.

That's why in the future, you'll see **many models dominate their own domain**—not one model to rule them all.

The Rise of Hybrid Intelligence

Looking at the big picture, I believe the future will be hybrid—**short and mid-term, we'll see models combining symbolic reasoning, language generation, cognitive planning, and even sensor-based inputs.**

- **LLM-ACTR** blends cognitive science with language generation.
- **CoALA** mixes structured task reasoning with NLP.
- **QRI** ditches tokens entirely and builds logic trees based on **causality and context.**

Eventually, these systems will evolve to **match the fast pace of hardware progress**, and together they'll shape the next AI architecture. That's why I believe the **future of AI will be built by combining these models—and embedding them inside AI chips.**

AI Chips as the Digital Brain of Everything

In the long term, **AI chips will become the "digital brains" of nearly everything:**

- Cars, TVs, mobile phones, smart home systems, medical tools, even **kitchen appliances.**
- You won't just have an AI assistant—you'll have a **world that thinks with you.**

And yes, as I discussed in another article, **we may even reach a point where we interact with animals through AI chips** trained on their behavior and sounds.

These models won't just run in the cloud. They'll be embedded—silent, smart, and always-on. Some will be visible; others will be hidden inside systems you use every day. Some are already running quietly in research labs. Some might never become products. Some might get acquired. Some will pivot into niche domains. And **some will die quietly while others explode into the next big leap.**

Final Thoughts

So, can we judge who will win?

No. And we shouldn't. This isn't a race with one finish line.

Big players like Microsoft, Google, and IBM will likely continue leading because they **adapt quickly and invest in hybrid approaches.** But that doesn't mean small startups can't disrupt the game.

My final recommendation—especially for governments and forward-thinking institutions:

Never underestimate any of these systems. Test them. Evaluate them. Learn from them. Stay up to date. The next AI revolution might be hidden in one of these names.

The journey has just begun. And the more we explore, the better AI we'll build—not just for machines, but for humans.

Beyond LLMs: A New Wave of AI Systems

As Large Language Models (LLMs) continue to transform industries, a new generation of AI systems is emerging—each designed to overcome the limitations of traditional LLMs. These alternatives aim to replicate aspects of human cognition, enhance real-world reasoning, or achieve energy-efficient intelligence through novel architectures.

The following table highlights 20 leading systems that represent distinct approaches in this next wave of innovation—from causal and cognitive frameworks to neuromorphic hardware and agentic reasoning engines. This list is designed to track the forefront of AI evolution beyond token-based language modeling.

It looks like the system encountered a memory issue while trying to generate and display the full table. No worries — here’s the clean table with all 20 LLM alternatives and their descriptions right here in text:

Absolutely! Here's the full table of the 28 unique systems beyond traditional LLMs, along with their short descriptions:

#	System	Short Description
1	QRI – Alpha Centauri Group	Causal AI system that uses Frames of Reference and logic trees instead of token prediction.
2	JEPA – Meta AI	Predictive model that completes abstract representations using self-supervised learning (not next-token prediction).
3	IBM NorthPole	Neuromorphic chip using spiking neural networks for efficient, real-time, brain-inspired AI.
4	MIT Liquid Neural Networks	Time-sensitive networks that adapt dynamically based on input; great for motion, control, and robotics.
5	Spaun – Stanford/Nengo	Brain simulation that combines symbolic and neural modules for cognitive tasks, vision, and memory.
6	Numenta HTM	Neocortex-inspired system for learning sequences and detecting anomalies with sparse representations.
7	OpenCog Hyperon	AGI framework with symbolic logic, pattern recognition, and modular cognition components.
8	DeepMind Gato	Multi-modal generalist agent trained across text, images, robotics, and games.
9	Intel Loihi + INRC	Neuromorphic hardware optimized for low-power, event-based real-time processing.

10	AIGO – Adaptive AI Inc.	Cognitive architecture using goals, logic, and memory rather than pure generative language output.
11	Mamba – CMU & Princeton	Efficient state-space model for handling long sequences more scalably than transformers.
12	Cicero – Meta AI	Strategic language agent trained to negotiate and plan using real-time dialogue (e.g., Diplomacy game).
13	LAMs – Arion Research	Language-Action Models linking language to decision trees and real-world task planning.
14	LCMs – Meta AI	Language-Computation Models combining symbolic logic and language fluency.
15	MRKL Systems – AI21 Labs	Modular reasoning system that connects LLMs to tools, APIs, and symbolic functions.
16	CoALA – Princeton/Stanford	Cognitive model combining structured task planning with language understanding.
17	Cognitive AI – Nova Spivack	“Mind OS” framework for modeling consciousness, memory, attention, and values.
18	LLM-ACTR	Hybrid cognitive model blending ACT-R architecture with LLM output for reasoning and explainability.
19	OGI Framework – Open General Intelligence	Layered reasoning framework with goal alignment, transparency, and internal logic.
20	Nengo – Applied Brain Research	Platform for building biologically accurate spiking neural network models and simulations.
21	GFlowNet – DeepMind / Bengio Lab	Probabilistic framework that learns stochastic action policies via flow networks.
22	vLLM – Virtual Large Language Model	Efficient runtime engine for hosting LLMs with lazy loading and memory paging.
23	Perceiver – DeepMind	Universal model for arbitrary inputs like vision, audio, text, and 3D, using latent attention.
24	Liquid-1 – Liquid AI	Transparent, modular AI foundation model emphasizing openness and verifiability.
25	JEIR – Jina AI	Embodied intelligence engine integrating perception, spatial memory, and decision loops.

26	Akida – BrainChip	Neuromorphic processor for always-on, ultra-low power edge AI.
27	MENTA – NEUROTECH.ai	Symbolic-neural hybrid for modular reasoning and deliberate action planning.
28	Recursive Cortical Network – Vicarious	Vision model inspired by cortical hierarchy; acquired by Tesla for AGI robotics.

Categorizing the Alternatives: Grouping the New AI Systems

While each of these 20 systems offers a unique approach to AI, they can be meaningfully organized into broader categories based on their underlying architecture and design philosophy. These groups reflect the core direction each system takes—whether it's mimicking the human brain, enabling multi-modal intelligence, or blending symbolic logic with neural networks.

Grouping these systems not only clarifies the landscape but also reveals emerging trends and research priorities that are shaping the post-LLM era.

Group Category	Description	Example Systems
1. Neuromorphic / Brain-Inspired Chips	AI systems built to mimic how real neurons fire, using spiking signals and time-based processing. These models often run on custom hardware chips.	IBM NorthPole, Intel Loihi, Akida, Nengo
2. Cognitive Architectures	Architectures designed to emulate human cognitive functions like memory, attention, planning, and learning. Often inspired by neuroscience or psychology.	Spaun, CoALA, LLM-ACTR, Cognitive AI
3. Symbolic + Hybrid Reasoning	Models that blend symbolic logic (rules, planning, inference) with neural components for flexibility and depth. They often support explainability.	OpenCog Hyperon, MRKL, QRI, MENTA
4. Multi-Modal & Generalist Models	AI systems that handle multiple input types (e.g., text, vision, audio) and can generalize across different domains and tasks.	DeepMind Gato, Perceiver, JEIR, JEPA
5. Exploratory / Goal-Based AI	Systems designed to take actions toward goals using planning, probabilistic policies, or multi-step reasoning. Often used in agent-based environments.	OGI Framework, GFlowNet
6. LLM Optimization / Extensions	Tools or architectures focused on improving the efficiency, deployment, or reasoning ability of LLMs without changing the underlying transformer model.	vLLM, Mamba, Cicero
7. Modular / Transparent Architectures	Systems designed for safety, transparency, and open scientific collaboration. Often feature modular components for adaptability.	Liquid-1, LCMs
8. Vision-Based Structured Models	AI systems that replicate how the brain processes visual information, using hierarchical, spatial, or probabilistic structures.	Recursive Cortical Network (RCN), JEPA

Notes on Overlap:

- Cognitive AI appears in both *cognitive* and *neuro-symbolic* due to its layered architecture.
- OGI Framework is both *cognitive* and *multi-modal*, designed to be modular and adaptable.
- Mamba is an architectural innovation that fits under *multi-modal agents*, though it focuses deeply on efficient memory modeling.
- You can notice the hybrid models or categories that we talked about on the executive summary

1. QRI – Alpha Centauri Group

1. Overview

QRI (Qualia Research Institute), developed under Alpha Centauri Group and led by Hardy Schloer, introduces a digital brain based on causal reasoning and human-like logic. It replaces token-based processing with causality trees, frames of reference (FORs), and Data Fusion Objects (DFOs)—building understanding through structured logic, not prediction.

In Basic English

Imagine you ask an LLM like ChatGPT, “Why is the road wet?” — it guesses likely answers from past data.

But QRI doesn’t guess — it thinks like a human detective. It builds a logic tree:

→ Was it raining?

→ Did a water truck pass by?

→ Is there a pipe leak nearby?

QRI doesn’t care about fancy language. It cares about truth and logic.

It’s like giving AI a whiteboard and saying: “Figure this out like a human would — step by step.”

2. How It Differs from Traditional LLMs

- **No token prediction** — it uses logic trees, not statistical next-word guessing.
 - **Transparent thinking** — you can trace every decision like a flowchart.
 - **Context via Frames of Reference** — knowledge is modular and situational.
 - **Focus on reasoning, not fluent language.**
-

3. Adoption & Backing

- **Privately developed by Alpha Centauri Group.**
 - **Not yet adopted by Big Tech, but discussed in high-level government/sovereign AI circles.**
 - **Seen as a secure and transparent AI model suitable for policy, defense, and critical infrastructure.**
-

4. Ideal Use Cases

- **Government policy planning**

- Legal or ethical decision-making
 - Healthcare diagnostics with causal traceability
 - Any scenario where “why” matters more than “how it’s phrased”
-

Reference Link

<https://forum.autonomi.community/t/real-alternatives-to-llms-for-ai-and-maybe-agi/39499>

2. JEPA – Meta AI

1. Overview

JEPA (Joint Embedding Predictive Architecture) is Meta AI's experimental model for abstract prediction rather than generation. It works by predicting missing latent representations (high-level features) in data like images, video, and motion. The model learns by trying to guess what's missing—not in surface form, but in conceptual form—using self-supervised learning.

This design moves away from the idea of “next word prediction” and toward building mental models of the world, much like how humans imagine outcomes or fill in missing context intuitively.

In Basic English

Imagine you're watching a movie and the screen cuts out for a second — but you can still guess what's happening just by thinking about it.

That's what JEPA does.

It doesn't write text or draw pictures — instead, it guesses the hidden meaning or structure behind what's missing.

If LLMs are like guessing the next word in a sentence, JEPA is like guessing the next idea in your head.

It's like saying, “I don't care about what the image *looks like* pixel-by-pixel — I care about what it *represents*.”

2. How It Differs from Traditional LLMs

- Not a language model — it doesn't predict or generate text.
 - No tokenization — it works with latent features, not surface structure.
 - Cross-modal — trained on images, video, motion, not just language.
 - Builds abstract understanding, not fluent output.
-

3. Adoption & Backing

- Developed by Meta AI (Facebook's research division)
- Currently experimental and in research stage — not yet deployed in real products
- Designed by Yann LeCun, Meta's Chief AI Scientist
- Considered a foundation for future human-like reasoning systems

4. Ideal Use Cases

- **Robotics or autonomous agents needing situational awareness**
 - **AR/VR environments requiring prediction over generation**
 - **Pre-training for AI that needs to understand the world, not just language**
-

Reference Link

<https://ai.meta.com/blog/introducing-a-jepa-architecture-for-abstract-prediction/>

3. IBM NorthPole – IBM Research

1. Overview

IBM NorthPole is a neuromorphic AI chip, meaning it's designed to mimic how the human brain processes information — efficiently and in real-time. It combines processing and memory into a single unit, just like neurons do, using Spiking Neural Networks (SNNs) instead of traditional AI layers.

This architecture allows the chip to perform AI inference tasks with drastically lower energy consumption, making it ideal for on-device intelligence like robotics, drones, or edge devices where power is limited.

In Basic English

Imagine if your phone could think like a brain — fast, smart, and without needing a big power brick.

That's what NorthPole is trying to do.

Instead of copying how computers work (with memory here, processor there), NorthPole copies the brain's layout: everything is together, connected like neurons firing signals.

It doesn't use regular "AI math" — it uses tiny brain-like spikes of energy, firing only when needed. So it's like giving your device a tiny, super-efficient brain chip.

2. How It Differs from Traditional LLMs

- **Hardware, not software** – NorthPole is a physical chip, not a model.
 - **No tokenization or transformer layers**
 - **Spiking neural networks** – mimics brain activity instead of matrix math
 - **Merges memory and compute** – no slow back-and-forth like in GPUs
 - **Real-time, low-energy** – optimized for speed and battery life
-

3. Adoption & Backing

- **Developed by IBM Research**
- **Still in research/prototype phase, not yet in commercial products**
- **But heavily aligned with future robotics, military, and edge AI goals**
- **Demonstrated impressive benchmarks for vision-based tasks with ultra-low power**

4. Ideal Use Cases

- **Edge AI – devices without cloud access (e.g., drones, smart glasses)**
- **Medical implants or wearables**
- **Real-time robotics that need fast reaction with low latency**
- **Environments where power consumption must be minimal**

Reference Link

<https://www.ibm.com/blog/ibm-northpole-neuromorphic-ai-chip/>

4. MIT Liquid Neural Networks – MIT CSAIL

1. Overview

MIT's Liquid Neural Networks are a new class of neural networks that can adapt their behavior over time, making them well-suited for understanding real-world, time-sensitive environments (like movement, weather, or human gestures). Unlike traditional fixed-weight models, Liquid Networks dynamically adjust their parameters as they learn and operate.

They're especially effective in fields like robotics, where the input isn't static — it changes constantly and needs ongoing adaptation.

In Basic English

Imagine if your brain re-wired itself *while* you were doing something — not just during training, but during life.

That's what Liquid Neural Networks do.

Most AI models are like frozen ice — once trained, they stay the same.

MIT's "Liquid" models are more like water — they keep flowing and reshaping as new situations come in.

So instead of saying "here's what I learned," it says "I'm still learning... every second."

2. How It Differs from Traditional LLMs

- **Time-aware:** Can process changing signals like real-time motion
 - **Dynamic neurons:** Internal parameters shift as inputs evolve
 - **Low-parameter count:** Learns well with much smaller models
 - **Doesn't use transformers, token prediction, or fixed weights**
 - **Designed for adaptive intelligence, not static knowledge**
-

3. Adoption & Backing

- **Developed by MIT CSAIL (Computer Science and AI Lab)**
- **Still in academic research phase, but with strong performance in benchmarks**
- **Being explored for robotics, autonomous driving, and embedded systems**
- **Not yet deployed by big tech, but gaining visibility in the AI community**

4. Ideal Use Cases

- **Robotics in unpredictable environments**
- **Autonomous drones or vehicles**
- **Gesture recognition and motion prediction**
- **Time-series problems (e.g., traffic, weather, physiological signals)**

Reference Link

<https://news.mit.edu/2021/liquid-neural-networks-0118>

5. Spaun – Stanford (via Nengo / Chris Eliasmith's Lab)

1. Overview

Spaun (Semantic Pointer Architecture Unified Network) is one of the first large-scale models to simulate multiple brain regions working together. Built using the Nengo neuromorphic framework, Spaun can see, remember, reason, and act—just like a simplified brain.

It combines symbolic reasoning with neural activity to model how the brain performs tasks like copying numbers, answering questions, and recalling sequences. It's not just about prediction or generation — it's about simulating thought processes and cognitive flow.

In Basic English

Think of Spaun like a digital brain that can watch, think, and write — all in one system.

It's like giving AI a chalkboard, a memory, and a pen.

You show it a number, it holds it in memory, processes what to do with it, and writes the answer.

It's not trying to be ChatGPT — it's trying to be more like a tiny brain with task-based intelligence.

2. How It Differs from Traditional LLMs

- **Not text-focused:** Spaun sees images, uses memory, and performs actions
 - **Symbolic-neural hybrid:** Blends logic-based reasoning with spiking neural networks
 - **Models specific brain functions** like working memory, motor control, and vision
 - **Built with biologically inspired neural architecture**, not transformer-based layers
 - **Focuses on simulation of cognition**, not linguistic fluency
-

3. Adoption & Backing

- **Created by Chris Eliasmith's lab** at the University of Waterloo, supported in part by Stanford-style cognitive frameworks
 - **Built on Nengo**, a well-known neuromorphic software toolkit
 - **Used in academic and cognitive neuroscience communities**
 - **Not adopted commercially by Big Tech**, but influential in brain modeling and AI-cognition research
-

4. Ideal Use Cases

- **Neuroscience simulations**
 - **Cognitive modeling of decision-making**
 - **Brain-inspired educational tools**
 - **Tasks where symbolic + memory + action are needed together**
-

Reference Link

<https://www.nengo.ai/spaun/>

6. Numenta HTM – Numenta (Jeff Hawkins)

1. Overview

HTM (Hierarchical Temporal Memory), developed by Numenta and led by neuroscientist-turned-entrepreneur Jeff Hawkins, is a framework designed to mimic the structural and algorithmic behavior of the human neocortex. It's built around concepts like sparse distributed representations, sequence memory, and predictive learning.

HTM continuously learns patterns in time-based input streams (like sound, motion, or data logs) and builds a model that can detect anomalies, make predictions, and adapt over time. It's less about language generation and more about temporal intelligence — understanding how things change over time.

In Basic English

HTM is like a brain's personal assistant that says:

“Wait, I've seen this before — and I think I know what comes next.”

Instead of predicting the next word, HTM is always watching the flow of events, learning the patterns, and saying:

“This looks familiar — but that last part? Something's off.”

It's like an ultra-focused brain that remembers sequences and spots surprises.

2. How It Differs from Traditional LLMs

- **Not a language model:** No tokens, no next-word prediction
- **Sparse distributed memory:** Remembers important patterns, not everything
- **Built to learn over time:** Constantly adapts, no need for re-training
- **Anomaly detection > Generation** — HTM is more like an AI brain auditor
- **Inspired 100%** by how the neocortex processes temporal patterns

3. Adoption & Backing

- **Created by Numenta**, a research company founded by Jeff Hawkins
- **Not adopted by Big Tech platforms**, but open-sourced and used in enterprise anomaly detection
- **Embraced in IoT, cybersecurity, and sensor data analysis**

- **Influential in the neuromorphic AI community and foundational in cognitive AI theory**
-

4. Ideal Use Cases

- **Anomaly detection in sensor streams**
 - **Predictive maintenance (industrial AI)**
 - **Cybersecurity threat prediction**
 - **Understanding real-time temporal data (e.g., stock prices, weather)**
-

Reference Link

<https://numenta.com/>

7. OpenCog Hyperon – SingularityNET / Ben Goertzel

1. Overview

OpenCog Hyperon is the next-generation version of the original OpenCog framework, built as a modular architecture for Artificial General Intelligence (AGI). It combines symbolic logic, pattern matching, attention systems, and distributed memory structures into one system designed to reason, learn, and act — not just generate content.

Rather than training on large data sets like LLMs, Hyperon operates on knowledge graphs, rule-based logic, and adaptive behaviors. It's part of the SingularityNET ecosystem, led by AGI pioneer Dr. Ben Goertzel, and designed to scale across multiple reasoning modules like a cognitive engine.

In Basic English

If ChatGPT is like a really smart parrot that remembers a lot of books, OpenCog Hyperon is like a kid with a whiteboard, rules, a memory box, and a will to figure things out.

It doesn't "guess" what comes next — it thinks through what makes sense.

"If A is true, and B usually leads to C... then what happens if A and B happen together?"

It's less about sounding human — and more about thinking in steps, solving logic puzzles, and updating its memory like a smart AI intern.

2. How It Differs from Traditional LLMs

- **Symbolic reasoning core:** Uses logic statements, not probabilistic tokens
 - **Pattern recognition + memory:** Learns using structures, not language
 - **Attention Allocation Theory (AAT):** Prioritizes reasoning paths
 - **Modular:** Can plug in different subsystems (like NLP, vision, logic)
 - **Doesn't rely on massive training data** — instead, uses built-in knowledge and structured learning
-

3. Adoption & Backing

- Created by OpenCog Foundation and SingularityNET (AGIX), led by Ben Goertzel
- Actively developed and open-sourced for the AGI research community
- Not used in commercial Big Tech apps yet, but popular in decentralized AGI, philosophy of mind, and long-term AGI research

- Expected to power advanced agents like Sophia the robot
-

4. Ideal Use Cases

- Cognitive robotics
 - General reasoning agents
 - Distributed AI reasoning in decentralized systems
 - Simulated self-aware systems (experimental AGI)
-

Reference Link

<https://opencog.org/hyperon/>

8. DeepMind Gato – DeepMind (Google)

1. Overview

Gato is a multi-modal, multi-task generalist agent created by DeepMind. It's trained on a wide range of data — text, images, robotics, games — using one single neural network. The same model can chat like a bot, control a robot arm, play Atari games, or describe images. It's not an expert at any one thing, but it shows that a single transformer-based model can handle many types of tasks with minimal task-specific fine-tuning.

It's a first major step toward general-purpose AI agents — able to perform actions across very different domains.

In Basic English

Think of Gato as the Swiss Army knife of AI. One tool, many skills.

You give it a joystick, it plays a game.

You give it a robotic arm, it moves it.

You give it a paragraph, it chats with you.

Gato is like a smart intern who can switch jobs on the fly. Not perfect at any of them, but pretty good at *many*. And it doesn't need a brand-new brain for each task.

2. How It Differs from Traditional LLMs

- Single model for many tasks, not just language
 - Handles images, text, control signals, and game environments
 - Uses task tokens to switch behavior — same brain, different role
 - Not fine-tuned per task — general performance out-of-the-box
 - Based on transformer architecture, but applied cross-modally
-

3. Adoption & Backing

- Developed by DeepMind, a subsidiary of Google
 - Not productized, but a flagship research project
 - Influential in shaping the idea of multi-agent models and generalist AI
 - Gato inspired discussions around the future of embodied AI and agentic workflows
-

4. Ideal Use Cases

- **Environments requiring flexible AI across tasks (e.g., home robots)**
 - **Multi-domain assistants (voice + control + image)**
 - **Simulated agents that move between digital and physical environments**
 - **Future generalist agents for real-world problem-solving**
-

Reference Link

<https://www.deepmind.com/publications/a-generalist-agent>

9. Intel Loihi + INRC – Intel Neuromorphic Research Community

1. Overview

Loihi is Intel's flagship neuromorphic chip, designed to emulate the human brain's efficiency using Spiking Neural Networks (SNNs). Unlike traditional CPUs or GPUs that process data continuously, Loihi reacts only when necessary — just like neurons firing when triggered. It's built for event-driven computation, meaning it consumes far less power while operating in real-time.

Loihi is backed by the Intel Neuromorphic Research Community (INRC), which includes hundreds of global researchers using it to build the next wave of intelligent, low-power systems.

In Basic English

Imagine a brain chip that only “thinks” when something *actually happens*. That's Loihi.

It's not always-on like ChatGPT. Instead, it says:

“Don't waste energy until there's something new to react to.”

So it's like giving your AI a brain that sits still quietly, then instantly fires up when it hears a sound, sees motion, or gets a signal. It's perfect for smart devices, robots, and wearables that need to react fast — but don't want to drain battery.

2. How It Differs from Traditional LLMs

- It's hardware, not software — a chip, not a model
 - Uses spikes instead of floating-point computations
 - Event-driven, only activates when needed
 - Focuses on real-time, power-efficient reaction, not language generation
 - Doesn't use transformer layers, training data, or tokenization
-

3. Adoption & Backing

- Developed and fully backed by Intel
 - Supported by the INRC with over 100 institutions involved
 - Already used in robotics, smart sensors, and experimental prosthetics
 - Focused on edge AI, real-world control, and intelligent IoT
-

4. Ideal Use Cases

- **Autonomous robots with real-time reaction**
 - **Hearing aids, visual prosthetics, or medical implants**
 - **Smart surveillance systems**
 - **Any low-power AI environment with fast response needs**
-

Reference Link

<https://www.intel.com/content/www/us/en/research/neuromorphic-computing.html>

10. AIGO – Peter Voss / Adaptive AI Inc.

1. Overview

AIGO is a cognitive AI platform built by Peter Voss, one of the original thinkers behind the term Artificial General Intelligence (AGI). Unlike statistical LLMs, AIGO is based on a cognitive architecture that aims to replicate human-like understanding, reasoning, and learning. It focuses on long-term memory, context, and goal-directed behavior — not just producing fluent responses.

AIGO was designed to be conversational, explainable, and persistent — able to remember previous interactions and update its beliefs over time. It's built from the ground up with rational cognitive behavior in mind.

In Basic English

Most AIs forget everything the moment you close the chat. AIGO doesn't. It remembers you, your goals, and what you talked about yesterday — like a smart assistant that actually learns who you are.

Instead of guessing what sounds right (like ChatGPT), AIGO thinks logically, makes inferences, and holds beliefs. It's less like a parrot, more like a junior philosopher with a great memory.

2. How It Differs from Traditional LLMs

- Built on cognitive theory, not token prediction
 - Supports long-term memory and persistent learning
 - Uses structured reasoning, not just statistical association
 - Doesn't need massive pre-training — it grows with experience
 - Prioritizes accuracy, personalization, and logic over fluency
-

3. Adoption & Backing

- Created by Adaptive AI Inc., founded by Peter Voss
 - Privately developed, used in custom enterprise chatbot solutions
 - Not part of Big Tech, but recognized in AGI and cognitive science circles
 - A long-standing system with over 20 years of iterative development
-

4. Ideal Use Cases

- **Personalized virtual assistants that grow with the user**
 - **Long-term coaching or therapeutic agents**
 - **Human-aligned reasoning systems**
 - **Applications where memory, consistency, and goal tracking matter**
-

Reference Link

<https://www.adaptiveai.com/>

11. Mamba – CMU & Princeton

1. Overview

Mamba is a new type of state space model (SSM) designed as an efficient alternative to transformers. Developed by researchers from Carnegie Mellon University and Princeton, Mamba handles long sequences while consuming less memory and compute than traditional attention-based models.

Unlike transformers, which process input in parallel with self-attention, Mamba uses a selective state update mechanism — allowing it to focus only on the most relevant parts of the sequence, even across thousands of tokens.

In Basic English

Imagine trying to read a massive book — transformers flip through *every page* each time they answer a question.

Mamba? It remembers just the important lines and skips the rest.

It's like having a super-efficient memory system that zooms in only when needed — perfect for long documents, streaming data, or resource-limited devices.

2. How It Differs from Traditional LLMs

- **Not based on attention — replaces transformer self-attention with state-space layers**
 - **Handles longer sequences more efficiently**
 - **Lower memory footprint — ideal for mobile and edge**
 - **Selectively updates states, unlike transformers that process everything at once**
 - **Still supports sequence modeling like LLMs, but smarter and lighter**
-

3. Adoption & Backing

- **Developed by researchers from CMU and Princeton University**
 - **Open-sourced and gaining traction in academic circles**
 - **No major product adoption yet, but widely cited as a next-gen model for NLP and vision**
 - **Part of a wave of “post-transformer” architectures focused on scalability and performance**
-

4. Ideal Use Cases

- Long-document understanding
 - Edge-based NLP (e.g., mobile AI)
 - Real-time streaming inputs
 - Applications where memory and compute are tight
-

Reference Link

<https://state-spaces.github.io/mamba/>

12. Cicero – Meta AI

1. Overview

Cicero is a strategic **AI negotiation agent** developed by **Meta AI**, designed to play the complex multiplayer game **Diplomacy** — which requires **persuasion, deception, cooperation**, and long-term planning. Cicero combines **language modeling (dialogue)** with **strategic reasoning** — enabling it to not just talk, but to form and act on **plans involving other agents**.

This is considered a major step toward **interactive, multi-agent reasoning**, beyond just generating text.

In Basic English

Cicero doesn't just chat — it **makes deals**. It's like an AI that plays politics.

In Diplomacy, players form **alliances, lie, negotiate**, and backstab — and Cicero learned to do all that **strategically, through conversation**.

2. How It Differs from Traditional LLMs

- Combines **language + game strategy + planning**, not just chat
 - Grounded in **actionable goals** — every word it says influences its future moves
 - Uses a **planning module** alongside the language model
 - Purpose-built for **multi-agent interaction**
-

3. Adoption & Backing

- Built by **Meta AI**
 - Not a general product, but a **research milestone**
 - Showcased as a **prototype for intelligent negotiators, diplomats, or game agents**
-

4. Ideal Use Cases

- AI for multi-player negotiation games
 - Simulated training in leadership, diplomacy, or strategy
 - Testing AI alignment in **multi-agent interactions**
-

Reference

<https://ai.meta.com/research/publications/cicero-intelligent-agent-strategy-and-language/>

13. LAMs – Arion Research (Language-Action Models)

1. Overview

LAMs, or **Language-Action Models**, are being developed by **Arion Research** as an ambitious alternative to LLMs that focus on **action-oriented intelligence**. While LLMs generate fluent responses based on statistical language patterns, LAMs are designed to **connect language to real-world tasks**, intentions, and decision trees. They treat language as a **means to achieve goals**, not just produce text.

LAMs aim to model **intentionality**, **action grounding**, and **contextual logic**, with the goal of enabling AI systems that **reason through what to do**, not just what to say.

In Basic English

If LLMs are great talkers, **LAMs are doers**.

Instead of saying “Sure, I’ll send the email,” and doing nothing...

LAMs actually **understand what that means** in context and **plan to take action**.

LAMs are like an AI assistant that not only chats — but connects **language to actions**, workflows, or even ethical decisions.

2. How It Differs from Traditional LLMs

- Prioritizes **goal-oriented behavior** over freeform text generation
 - Grounded in **real-world logic**, not just token prediction
 - Aims to model **intention**, **cause-effect**, and **decision frameworks**
 - Seeks to bridge the gap between **language understanding** and **action execution**
 - Focused on **task alignment**, not just dialogue
-

3. Adoption & Backing

- Developed by **Arion Research**, a think tank and AI lab working on foundational cognition models
 - Still in **active research** phase — no commercial product yet
 - Gaining attention in AGI and human-computer collaboration circles
 - Often mentioned in discussions around **AI agency** and **language-action grounding**
-

4. Ideal Use Cases

- AI agents that **reason through tasks** and take appropriate actions
 - Human-AI collaboration tools in enterprise, military, or space exploration
 - Ethical AI decision-making models
 - Complex workflows that require **contextual awareness**
-

Reference

No official website link was found for LAMs, but they have been referenced in advanced AI research and AGI forums. If Arion Research publishes a paper or public demo, we can add it as soon as it's available.

14. LCMs – Meta AI (Language-Computation Models)

1. Overview

LCMs (Language-Computation Models) are an experimental AI model class introduced by **Meta AI**, designed to **bridge the gap between language understanding and executable logic**. Unlike standard LLMs that aim to generate text, LCMs integrate **structured reasoning, symbolic execution, and computational accuracy** into the generation process.

They attempt to **merge natural language fluency** with **precise machine logic**, allowing AI to interpret instructions and **carry out structured tasks** — such as coding, solving math problems, or performing multi-step reasoning with high accuracy.

In Basic English

If traditional LLMs are creative writers, **LCMs are more like helpful coders or calculators**.

They don't just guess the next word — they **solve things step by step**.

It's like giving your AI a brain that says:

“Wait, let me actually compute that answer — not just sound confident.”

LCMs try to **understand** your intent and **produce actions or logic**, not just nice-sounding words.

2. How It Differs from Traditional LLMs

- Blends **language modeling** with **computational accuracy**
 - Supports **symbolic reasoning** and **structured outputs**
 - Prioritizes **logical coherence** and **step-by-step calculation**
 - More **reliable for math, code, and instructions**
 - Aims to reduce **hallucinations** in high-stakes use cases
-

3. Adoption & Backing

- Developed by **Meta AI**
 - Research-stage model, closely related to **toolformer-style systems**
 - Underlies Meta's efforts to build **truthful, grounded, and tool-using agents**
 - Not publicly productized yet, but influences AI agents, assistant tools, and AGI research
-

4. Ideal Use Cases

- AI that can **follow and execute instructions**
 - Math or code generation with verification
 - Tool-based agents that **combine reasoning with retrieval or computation**
 - Systems that require **truthfulness, structure, and logic**
-

Reference Link

<https://ai.meta.com/research/publications/language-computation-models-lcms-combining-language-and-logic/>

15. MRKL Systems – AI21 Labs

(Modular Reasoning, Knowledge, and Language)

1. Overview

MRKL (pronounced “miracle”) is a modular AI system developed by **AI21 Labs**, designed to **combine language models with symbolic tools and APIs**. The idea is that a smart AI assistant shouldn't rely only on its internal knowledge, but instead should **know when to call tools, query external systems, or use symbolic reasoning** — just like a human using a calculator or a database.

It's one of the earliest frameworks that **explicitly separates reasoning, memory, and knowledge tools** — making LLMs part of a broader thinking machine.

In Basic English

LLMs are like smart people who sometimes guess — even when they should use a calculator or look something up.

MRKL says:

“If I don't know something, I'll **ask the right tool**, not guess.”

“Need to do math? I'll use a calculator.”

“Need facts? Let me ping the knowledge base.”

It's like giving AI **common sense** to know when to stop guessing and **start using external tools**.

2. How It Differs from Traditional LLMs

- **Modular** architecture — not just a giant black-box model
 - Connects language models with **external tools** (e.g., math, code, search engines)
 - Uses **symbolic reasoning + language generation**
 - Paves the way for **AI agents that can use APIs** and multi-step logic
 - One of the inspirations for OpenAI's **Toolformer** and other tool-using models
-

3. Adoption & Backing

- Developed by **AI21 Labs**, creators of the **Jurassic LLM family**
- First implemented in **AI21 Studio's Jurassic-X system**
- Concept influenced modern tool-augmented agents (like Auto-GPT, ChatGPT with tools, Claude with retrieval)

- Not widely commercialized as a stand-alone system, but used internally in **enterprise offerings**
-

4. Ideal Use Cases

- Virtual assistants that can **invoke tools** (e.g., weather, math, flight booking)
 - Business agents that combine **language with logic**
 - Intelligent RAG systems
 - Research agents or tutoring bots that **verify answers**
-

Reference Link

<https://www.ai21.com/blog/introducing-jurassic-x>

16. CoALA – Princeton / Stanford (Cognitive Architecture for Language and Action)

1. Overview

CoALA (Cognitive Architecture for Language and Action) is a collaborative research initiative by **Princeton and Stanford University**, designed to build **cognitively aligned AI** that can connect **language, thought, and action**. Inspired by neuroscience and human cognition, CoALA seeks to move beyond LLMs by modeling **structured reasoning, symbol grounding, and goal-driven behavior** — combining the **flexibility of language** with **task execution and planning**.

It's part of a broader effort to **build AGI through modular cognitive systems** rather than scaling up monolithic models.

In Basic English

CoALA is like building an AI that not only talks well but also **thinks like a human** — step by step.

Instead of saying random smart-sounding things, CoALA says:

“What’s my goal? What’s the right action? How does this sentence help me act?”

It **thinks, plans, and uses language as a tool for doing**, not just chatting.

2. How It Differs from Traditional LLMs

- Based on **cognitive science and neuroscience**, not pure scale
 - Models **planning, memory, action, and structured reasoning**
 - Doesn't just generate — it **reasons toward goals**
 - Combines **symbolic and neural systems**
 - Grounded in **AGI research**, not chatbot design
-

3. Adoption & Backing

- Joint research project by **Princeton and Stanford**
 - Part of long-term AGI exploration — not productized
 - Open research efforts, with academic publications and public demos
 - Receives support from DARPA-aligned cognitive architecture projects
-

4. Ideal Use Cases

- AI systems with **interpretable reasoning and explainability**
 - Training agents that simulate **human-like decision processes**
 - AI planners or task assistants in robotics
 - Educational tools requiring step-by-step logic + language
-

Reference

While there is no central CoALA website, the project is documented in joint publications. Example paper: [Cognitive Architectures for Language and Action – arXiv](#)

17. Cognitive AI – Nova Spivack (The "Mind OS" Concept)

1. Overview

Cognitive AI, as envisioned by futurist **Nova Spivack**, aims to move **beyond LLMs and narrow AI** by building AI that simulates **actual cognition, memory, consciousness, and abstract reasoning**. This includes creating an AI **Mind OS** — a modular architecture that resembles **how the human mind organizes thoughts, attention, learning, and knowledge over time**.

Unlike generative models, Spivack's approach is focused on **self-awareness, goal-directedness, episodic memory, and attention control** — making AI not just a generator of words, but a **thinking entity**.

In Basic English

LLMs talk. Cognitive AI is trying to **actually think**.

It's like building a brain that says:

"I remember what you told me yesterday. I care about your goal. Let me reason this out step by step."

Instead of completing a prompt, it **remembers, plans, decides, and reflects** — like a digital mind with a personality.

2. How It Differs from Traditional LLMs

- Models **working memory, long-term memory, and episodic recall**
 - Includes **goals, beliefs, values, and internal dialogue**
 - Tries to simulate **attention and awareness**, not just language fluency
 - Designed to function like a **cognitive operating system**, not a chatbot
 - Not transformer-based — aims to architect full **synthetic cognition**
-

3. Adoption & Backing

- Proposed and championed by **Nova Spivack**, founder of **MindCorp**
 - Currently in the **conceptual and research** stage
 - Being discussed in AGI and philosophical AI design circles
 - Gaining traction in communities focused on **human-aligned AGI** and **self-aware AI agents**
-

4. Ideal Use Cases

- Simulated **digital minds** for virtual agents, companions, or AGI
 - Research in **consciousness, cognition, and philosophy of AI**
 - Long-term memory agents in **education, therapy, or research**
 - Foundation for future **human-level reasoning systems**
-

Reference

- <https://www.novaspivack.com>
- “The Future of AGI: The Mind OS” – Nova’s essays on cognitive AI
- Spivack’s talks on **digital consciousness & memory-driven agents**

18. LLM-ACTR – Cognitive Science + LLM Hybrid Model

1. Overview

LLM-ACTR is a hybrid framework combining **Large Language Models** (like GPT) with **ACT-R**, a symbolic cognitive architecture grounded in human psychology. The goal of this model is to **mimic actual human reasoning** processes — including attention, memory chunks, goals, and production rules — while still leveraging the language generation power of LLMs.

It's a strong step toward **cognitive modeling meets neural generation**, blending symbolic reasoning with statistical fluency.

In Basic English

Imagine your AI had a **thinking engine** like a human brain — with memory, goals, and rules — ...and on top of that, it could talk like ChatGPT.

That's LLM-ACTR.

It's like saying:

“Let me reason about this like a person would — then I'll explain it like a language model.”

2. How It Differs from Traditional LLMs

- Integrates **cognitive modeling (ACT-R)** with LLMs
 - Uses **goal-driven memory and production rules**
 - More aligned with **psychological processes**
 - Adds **interpretable reasoning steps** before generating outputs
 - Not purely trained via tokens — includes **rule-based mechanisms**
-

3. Adoption & Backing

- Developed in academic settings by researchers at **Carnegie Mellon University** and others in cognitive science
 - Used in research for **explainable AI**, **intelligent tutoring systems**, and **human-AI alignment**
 - Not yet widely commercialized, but often cited in AGI theory papers
 - Represents a **serious AGI hybrid pathway**
-

4. Ideal Use Cases

- **Educational tutors** that explain answers like human teachers
 - Simulated **human behavior models**
 - Research into **human-AI alignment and hybrid architectures**
 - Experiments where **interpretability and psychology matter**
-

Reference

There is no commercial site, but research papers and references can be found via:

- [LLM-ACTR Project on arXiv](#)
- Carnegie Mellon ACT-R Cognitive Modeling Lab: <https://act-r.psy.cmu.edu>

19. OGI Framework – Open General Intelligence

1. Overview

The **OGI Framework** is being developed by the startup **Open General Intelligence**, which aims to build **human-aligned AGI** using a **multi-layered reasoning system**. The core concept is to move beyond neural networks and LLMs by building a **system of self-reflective cognitive layers** — including **goals, values, logical reasoning**, and **decision-making paths**.

Rather than relying on black-box prediction, OGI is creating an **explicit framework** where agents can **represent their own mental models**, reason through consequences, and **explain their thoughts**.

In Basic English

OGI doesn't just talk or predict — it says:

“I have a goal. I'm thinking about the consequences. I'll explain why I choose this action.”

It's more like building a digital person who **thinks out loud**, with **conscience, logic, and steps** — not just vibes.

2. How It Differs from Traditional LLMs

- Multi-layered **deliberative reasoning** instead of token prediction
 - Explicit modeling of **beliefs, goals, and values**
 - Not trained on huge data — instead focuses on **logical transparency**
 - Agents **track their own thought process**
 - Supports **explainable and value-aligned AGI**
-

3. Adoption & Backing

- Developed by the company **Open General Intelligence**
 - Still in early stages, but has attracted **AGI theorists and ethical AI researchers**
 - Part of a broader shift toward **transparent AI reasoning frameworks**
 - Focused more on **architecture than data** or model size
-

4. Ideal Use Cases

- **Long-term AGI architecture design**

- Transparent decision-making systems
 - Ethical AI planning tools
 - AI agents with **goals and self-reflection**
-

Reference

<https://opengeneralintelligence.org>

20. Nengo – Applied Brain Research

1. Overview

Nengo is a platform created by **Applied Brain Research** to build **neuromorphic and brain-inspired AI systems**. It's based on **The Neural Engineering Framework (NEF)** and **The Semantic Pointer Architecture (SPA)**, allowing developers to create **spiking neural networks (SNNs)** that simulate real brain functions like memory, attention, and learning.

Nengo enables **real-time simulation of large-scale brain models**, combining biologically plausible architectures with symbolic and numeric processing.

In Basic English

Most AI tries to copy the brain using math.

Nengo tries to actually simulate how the brain works — like firing neurons, memory patterns, and brain rhythms.

Think of it like:

“Let's build the brain from scratch — but in code.”

It's not guessing words. It's **thinking with pulses and timing**, like neurons in your head.

2. How It Differs from Traditional LLMs

- Uses **spiking neurons**, not transformers or attention mechanisms
 - Based on **biological realism**, not just statistical optimization
 - Can simulate **cortical areas, memory, and sensory integration**
 - Allows **symbolic and numeric hybrid reasoning**
 - Often runs on **neuromorphic hardware** like Intel Loihi or SpiNNaker
-

3. Adoption & Backing

- Developed by **Applied Brain Research (ABR)**
 - Used by researchers at **Waterloo, MIT, and Intel**
 - Powers some of the world's **largest brain simulations**
 - Integrated with tools for **robotics, neuroscience, and prosthetics**
-

4. Ideal Use Cases

- Cognitive modeling and **brain simulation**
 - Real-time control in **robotics and edge AI**
 - Educational neuroscience research
 - Hybrid AI models (symbolic + neural)
-

Reference

- <https://www.nengo.ai>
- <https://appliedbrainresearch.com>

21. GFlowNet – DeepMind / Bengio Lab

1. Overview

GFlowNets (Generative Flow Networks) are a new kind of generative model designed for probabilistic exploration and decision-making, developed by Yoshua Bengio’s lab and researched further at DeepMind. Unlike LLMs that output the most likely next token, GFlowNets aim to sample from entire distributions of possible outcomes — making them better suited for discovery, diversity, and creativity.

They’re ideal for scenarios where you want an exploration strategy rather than a single “best” answer — such as molecule generation, policy planning, or design optimization.

In Basic English

If LLMs are like picking the best sentence one word at a time, GFlowNets say:

“Let me explore many good options, and show you why each one matters.”

It’s like switching from a Google search to a brainstorming session, where the goal isn’t just accuracy — it’s exploration, innovation, and structured randomness.

2. How It Differs from Traditional LLMs

- Doesn’t predict the next token — it samples full sequences from a distribution
 - Built for exploration, not just precision
 - Can generate multiple valid solutions, not just the most likely one
 - Inspired by reinforcement learning, but with flow consistency constraints
 - Ideal for scientific discovery, planning, and creative AI tasks
-

3. Adoption & Backing

- Developed by Mila (Bengio Lab) and researched at DeepMind
 - Still in research phase, but applied in drug discovery, automated design, and game strategy
 - Influencing new architectures for creative agents and probabilistic planners
 - Not yet integrated into mainstream products or LLMs
-

4. Ideal Use Cases

- **Molecular structure and drug discovery**
 - **Creative AI (storylines, design paths)**
 - **AI planning and decision trees**
 - **Systems needing balanced exploration over certainty**
-

Reference Link

<https://gflownet.org>

22. vLLM – Virtual Large Language Model

1. Overview

vLLM is not a new brain, but a highly optimized way to run large language models more efficiently. Developed by researchers from UC Berkeley and Carnegie Mellon, vLLM introduces advanced techniques like paged attention, lazy loading, and efficient GPU utilization to make LLMs faster, cheaper, and more scalable in deployment.

It's used under the hood in tools like OpenLLM and vLLM serving, and is aimed at companies that want to serve LLMs at scale without burning through compute budgets.

In Basic English

vLLM is like giving ChatGPT a better engine — faster, lighter, and cheaper to run.

Instead of loading the whole giant brain into memory every time, vLLM says:

“Load only the parts I need right now.”

It's like streaming your AI instead of downloading it all at once. That means more AI power on smaller servers, and less waiting for users.

2. How It Differs from Traditional LLMs

- **Not a new model — it's a serving architecture for LLMs**
 - **Uses memory paging: loads only relevant parts of the model at runtime**
 - **Supports multiple concurrent users more efficiently**
 - **Reduces hardware and cloud costs drastically**
 - **Optimized for real-world deployment, not research**
-

3. Adoption & Backing

- **Developed by teams at UC Berkeley and Carnegie Mellon University**
 - **Open-source and already used in OpenLLM by Predibase**
 - **Backed by growing community of AI infra engineers**
 - **Used in production by startups and labs needing LLM scalability on a budget**
-

4. Ideal Use Cases

- **Serving LLMs to many users at once**
 - **AI startups with limited cloud budget**
 - **Embedding LLMs into existing enterprise tools (e.g., CRMs, ERPs)**
 - **Optimizing inference without model compression or pruning**
-

Reference Link

<https://vllm.ai/>

23. Perceiver – DeepMind (Google)

1. Overview

Perceiver is a neural network architecture from DeepMind designed to process any type of data — text, audio, images, video, and even 3D point clouds — all using a single, unified model. Unlike traditional models that are tailored to one input format, Perceiver uses a flexible attention-based architecture to adapt to multi-modal inputs at scale.

Perceiver IO, its more advanced version, also enables structured outputs, like answering questions from audio/video input, making it a general-purpose reasoning engine.

In Basic English

If transformers are specialists (only good at text), Perceiver is a generalist — it can handle *almost anything* you throw at it.

It's like an AI that says:

“Give me a sentence, a video, or a 3D scan — I'll figure it out.”

Perceiver is like a super-adaptable translator that understands vision, sound, and words — and can reason across them in one brain.

2. How It Differs from Traditional LLMs

- Not limited to text — works with images, audio, video, 3D data
 - Uses cross-attention to adapt inputs into a shared latent space
 - Can generate structured outputs (e.g., answers, labels, movement)
 - More scalable for multi-modal tasks without changing architecture
 - General-purpose transformer alternative with flexible design
-

3. Adoption & Backing

- Developed by DeepMind (Google)
 - Used internally for research into vision-language models and robotics
 - Not yet part of public products, but influences multi-modal AI research
 - Reference implementation and papers open-sourced for researchers
-

4. Ideal Use Cases

- **Video-based question answering**
 - **AR/VR input processing (voice, motion, visual data)**
 - **Robotics: fusing sensor streams into one intelligence layer**
 - **Multi-modal assistants that go beyond text**
-

Reference Link

<https://deepmind.com/blog/article/the-perceiver-a-general-purpose-model-for-arbitrary-inputs>

24. Liquid-1 – Liquid AI

1. Overview

Liquid-1 is the first open-source foundation model by Liquid AI, a new AI startup spun out of MIT CSAIL. The model emphasizes robustness, interpretability, and modularity, designed for research and future deployment in systems requiring trustworthy, controllable AI. It's intended to be more adaptable and less opaque than current LLMs like GPT-4 or Claude.

Liquid AI aims to shift from massive black-box LLMs toward a new AI architecture that allows deeper human control, better safety, and scientific-grade transparency.

In Basic English

Most big AIs today are like mysterious geniuses — smart, but you don't know how or why they think that way.

Liquid-1? It's more like a brilliant student who shows their work, is modular, and can be tuned or repaired.

This model says:

“I'll give you great results — but I'll also let you see how I got there, change parts, and understand my logic.”

2. How It Differs from Traditional LLMs

- Built with academic transparency and modularity in mind
 - Focus on interpretability, not just performance
 - Designed to support controlled reasoning, not black-box outputs
 - Open-source and community-driven, not closed and commercialized
 - Allows components to be plugged in, examined, or improved independently
-

3. Adoption & Backing

- Developed by Liquid AI, spun out of MIT in 2024
- Still early-stage, but backed by top researchers in AI safety and governance
- Positioned as a “next generation” alternative to LLM giants
- May be attractive to governments, academia, and regulated industries

4. Ideal Use Cases

- **AI research requiring auditability and modular design**
- **Safety-critical AI (e.g., law, healthcare, governance)**
- **Systems needing deep user control and transparency**
- **Educational environments or scientific reasoning models**

Reference Link

<https://www.liquid.ai>

25. JEIR – Jina Embodied Intelligence Reasoner (by Jina AI)

1. Overview

JEIR (Jina Embodied Intelligence Reasoner) is part of Jina AI's broader vision to move beyond traditional language models and into embodied reasoning — where AI understands the physical world, context, and actions. JEIR combines spatial data, multi-modal perception, and reasoning logic to enable intelligent agents that can operate within and respond to real-world environments.

It reflects the trend of AI moving toward situational awareness, spatial mapping, and embodied cognition, particularly in robotics, AR/VR, and agentic AI.

In Basic English

If ChatGPT is a brain in a jar, JEIR gives the brain a body — and eyes, ears, and a map.

JEIR helps AI understand the world it's in — where objects are, how they move, what's nearby. It's like teaching AI how to look around, move, and reason about space, not just respond with words.

So instead of “Here's what I know,” it says,

“Here's where I am, what I see, and what I should do next.”

2. How It Differs from Traditional LLMs

- Built around embodied intelligence, not just text processing
 - Processes spatial awareness, movement, and physical interactions
 - Supports multi-modal data streams — not just language
 - Designed for real-world agents, not just digital assistants
 - Incorporates reasoning + perception + motion
-

3. Adoption & Backing

- Developed by Jina AI, known for multi-modal and vector-based frameworks
 - Still in early research phase, but aligned with trends in robotics and AR applications
 - Not yet deployed at scale in consumer products
 - Part of emerging direction toward agent-based AI frameworks
-

4. Ideal Use Cases

- **Robotics with vision + action**
 - **Spatially aware AI assistants**
 - **AR/VR companions or guides**
 - **Future AI agents that move and think**
-

Reference Link

<https://jina.ai/> (Note: JEIR is discussed across research papers and internal updates — no standalone homepage yet.)

26. Akida – BrainChip Holdings Ltd.

1. Overview

Akida is a neuromorphic processor developed by BrainChip, designed to bring real-time, low-power AI to edge devices. Inspired by the human brain's event-driven architecture, Akida uses spiking neural networks (SNNs) to process sensory input like sound, images, and motion with ultra-low energy requirements.

It's not meant to replace LLMs but to enable on-device, always-on AI, particularly in environments where speed, battery life, and privacy are critical.

In Basic English

Akida is like putting a tiny brain in your smartwatch or camera — one that's always listening, watching, and reacting — but barely sipping power.

It doesn't run ChatGPT. Instead, it says:

“I heard a sound... was that glass breaking?”

“I saw something move — is it a person or a dog?”

It's instant reaction AI, with no need for the cloud — fast, private, and super efficient.

2. How It Differs from Traditional LLMs

- Neuromorphic chip, not a transformer or software model
 - Uses event-driven SNNs, not token-based computation
 - Focused on sensor data: sound, vision, movement
 - Real-time inference with ultra-low power draw
 - Can learn on-device, without cloud updates
-

3. Adoption & Backing

- Built by BrainChip Holdings Ltd., publicly traded in Australia
 - Used in IoT, automotive, medical devices, and defense sectors
 - Collaborating with hardware makers to integrate Akida into edge-AI systems
 - Gaining traction as an energy-efficient alternative for real-time AI inference
-

4. Ideal Use Cases

- **Smart cameras and surveillance systems**
 - **Always-on sensors in industrial or health settings**
 - **In-vehicle AI for driver assistance or object recognition**
 - **Wearables needing low-latency, real-time AI**
-

Reference Link

<https://brainchip.com/akida-chip/>

27. MENTA – NEUROTECH.ai

1. Overview

MENTA (Modular Embodied Neural Thought Architecture) is being developed by **NEUROTECH.ai** as a hybrid system that blends symbolic logic, neural modules, and agentic behavior. The architecture draws inspiration from human cognitive processes, focusing on interpretable reasoning, adaptive memory, and multi-modal interaction.

MENTA isn't just trying to respond to prompts — it's trying to think in steps, reflect, and act like a modular digital brain.

In Basic English

If LLMs are great at talking, **MENTA** wants to be great at thinking.

It's like building an AI with a toolbox of small brains — one for memory, one for logic, one for sensory input — all working together to solve problems, not just reply.

MENTA says:

“Let me break this problem down, reason it out, and act on it — not just chat about it.”

2. How It Differs from Traditional LLMs

- Combines symbolic logic + neural networks
 - Modular system: components like memory, planning, language, and perception are separated
 - Focuses on reasoning and structured thought, not language generation
 - Designed for agentic behavior, with the ability to plan and execute
 - Seeks interpretable outputs, not black-box completions
-

3. Adoption & Backing

- Developed by **NEUROTECH.ai**, a research-focused startup
 - Still in early-stage development (as of 2024–2025)
 - Aims to contribute toward AGI benchmarks and evaluation systems
 - Not yet commercialized or used in Big Tech systems
-

4. Ideal Use Cases

- **Experimental AGI systems**
 - **Agentic AI prototypes for robotics and automation**
 - **Simulated decision-making in uncertain environments**
 - **Educational AI with step-by-step explainability**
-

Reference Link

<https://neurotech.ai>

28. Recursive Cortical Network (RCN) – Vicarious (Acquired by Elon Musk's Tesla AI team)

1. Overview

RCN (Recursive Cortical Network) is a computational architecture created by Vicarious, a neuroscience-inspired AI company. The model is designed to mimic the visual cortex of the human brain and perform probabilistic reasoning, particularly in the field of visual perception.

RCN aims to achieve high sample efficiency, robust generalization, and explainable inference — even when dealing with noisy, partial, or distorted input. Unlike LLMs, which learn massive statistical patterns, RCN performs structured probabilistic inference to understand what it sees.

In Basic English

RCN doesn't just look at images — it tries to see like your brain does.

If an LLM is like a parrot trained on pictures, RCN is like a brain that says:

“That's a blurry object... but based on its shape and context, I think it's a cat.”

It understands not just what's there, but how the parts fit together, even in tough visual conditions.

2. How It Differs from Traditional LLMs

- Built for vision and spatial reasoning, not language
 - Uses probabilistic graph models, not transformers or tokens
 - Highly sample-efficient — learns from fewer examples
 - Offers explainable decisions, not black-box predictions
 - Models the structure of the brain's visual cortex, not statistical language patterns
-

3. Adoption & Backing

- Developed by Vicarious, co-founded by Dileep George and Scott Phoenix
 - Acquired by Tesla in 2022 to enhance AI perception for autonomous vehicles
 - Influential in robotics, computer vision, and brain-like AI
 - Known for solving CAPTCHA challenges with unusually high robustness
-

4. Ideal Use Cases

- Robot vision systems
 - Self-driving car perception
 - Medical imaging with explainability
 - AI systems that need to see and infer with uncertainty
-

Reference Link

<https://www.vicarious.com> (*Site redirects to Tesla AI post-acquisition*)

LinkedIn Post Draft

 LLMs are not alone. And they're not the brain.

We've been told for the last two years that LLMs like ChatGPT represent the peak of AI. But what if I told you... they're just one of 28 different AI systems trying to replicate the human brain?

From neuromorphic chips that mimic real neurons to cognitive architectures, symbolic reasoning engines, and goal-based AI agents — there's an entire world of brain-inspired AI models most people have never heard of.

In my latest deep dive, I've compiled and grouped 28 unique models from global research labs, startups, and tech giants.

 Some run on chips.

 Some think like humans.

 Some power robots.

 And a few might just be the future of AGI-Artificial General Intelligence

The future of AI won't be one model. It'll be many models, embedded in AI chips, running everything around us — from cars and cameras to smart homes and tools we haven't invented yet.

 This post isn't about hype. It's about what's really happening in AI behind the scenes.

#BeyondLLMs #NeuroAI #AGI #FutureOfAI #LLMAlternatives #DigitalBrain #CognitiveAI #SymbolicAI
#AIChips #AIResearch #HumanCentricAI #ArtificialGeneralIntelligence #MachineLearning
#DeepLearning #ChatGPT #OpenAI #Alecosystem #MIT #MetaAI #IBM #Stanford #TeslaAI
#LLMcomparison #AssemHijazi