NeuroAI and Beyond

How Brain Science Can Fix AI - And How AI Can Unlock Brain Secrets

Simplified Version of the NSF Workshop Report

Original Authors: Jean-Marc Fellous, Gert Cauwenberghs, Cornelia Fermuller, Yulia Sandamirskaya & Terrence Sejnowski

UC San Diego | *University of Maryland* | *Zurich UAS* | *Salk Institute*

What is this paper?

In August 2025, 70 top scientists from neuroscience, AI, robotics, and engineering gathered at an NSF-funded workshop in Arlington, Virginia. Their mission: figure out how the brain and artificial intelligence can help each other. This simplified version makes their groundbreaking ideas accessible to everyone - no PhD required.

Original paper: arXiv:2601.19955 (53 pages)

Simplified by: @muhamedfazalps7

What's Inside

1. The Big Picture: Two Revolutions That Need Each Other

The Central Idea

Artificial intelligence has achieved remarkable things. It can write poetry, pass medical exams, and generate photorealistic images. But there's a problem: AI still can't do the things a two-year-old finds easy - like walking across a cluttered room, catching a ball, or understanding that when someone hides a toy under a blanket, the toy still exists.

Meanwhile, neuroscience has made incredible progress understanding how the brain works. We now have detailed maps of brain connections (connectomes), we can record from thousands of neurons simultaneously, and we understand much about how memories form and how we learn.

The key insight of this paper is simple but profound: **AI and neuroscience need each other**. AI can help neuroscientists analyze massive datasets and test theories. Neuroscience can show AI researchers how to build systems that are more efficient, flexible, and robust.

How AI and Brains Differ

To understand why this collaboration matters, consider how different AI and brains actually are:

Feature	Current AI (like ChatGPT)	Human Brain
Energy	Uses as much power as a small city	Uses 20 watts - like a dim light bulb
Learning	Learns once from massive data	Learns continuously throughout life
Memory	Forgets old things when learning new	Consolidates memories during sleep
Body	Exists only as software	Deeply connected to a physical body
Development	No childhood - born fully formed	Years of gradual development
Feedback	Mostly one-directional	Massive two-way feedback loops

The Moravec's Paradox

In 1988, computer scientist Hans Moravec noticed something strange: the things humans find hard (chess, calculus, logic) are easy for computers, while the things we find effortless (walking, recognizing faces, catching a ball) are incredibly hard for computers.

Nearly 40 years later, this paradox still holds. AI can pass medical licensing exams and write publishable prose, but the most capable robot you can buy for your home is still basically a vacuum cleaner that bumps into furniture.

This paper argues that the solution is to look at how brains actually work - not just their mathematical approximations, but their physical architecture, their development, and their deep connection to bodies.

The Three Big Problems

The workshop identified three fundamental problems with current AI:

1. AI cannot interact with the physical world. Most AI exists only as software processing text and images. It has no body, no senses, no ability to act in the real world.

2. AI learns inflexibly. When you teach ChatGPT something new, it often "forgets" what it already knew. Humans learn new things while preserving old memories.

3. AI is energy-hungry. Training a single large AI model can emit as much carbon as five cars in their lifetimes. Brains do far more with far less energy.

Key Takeaway: The brain is the gold standard for intelligent, efficient, robust computation. AI needs to learn from it - not just copy it, but understand the principles that make it work.

2. Giving AI a Body: Why Robots Need to Touch the World

What Is Embodiment?

An embodied system is one that has a physical body, can sense its environment, and can act upon it. Think of a dog: it has legs to move, eyes and ears to sense, muscles to act, and a brain that coordinates everything. The dog learns about the world by interacting with it - sniffing, touching, tasting, running.

Current AI is like a brain in a jar. It processes text and images but has never touched a doorknob, felt the weight of a glass, or learned that fire is hot by getting burned. This fundamental disconnect may be why AI lacks common sense.

Why Does Embodiment Matter?

When an agent can explore autonomously, interact with, and be modified by a complex environment, it develops an **interactive understanding** of that environment. This understanding provides meaningful rewards and punishments that allow it to generate flexible, adaptive behavior.

Consider how a baby learns about gravity. No one teaches the baby that things fall down. The baby drops toys, watches them fall, reaches for them, and eventually builds an intuitive understanding of how gravity works. This understanding is grounded in physical experience - not in reading textbooks about physics.

Large language models, by contrast, learn about gravity from text descriptions. They know that 'objects fall downward' because they've read it millions of times, but they don't have the deep, embodied understanding that comes from physical interaction.

The Praxicon: A Dictionary of Actions

One fascinating idea from the paper is the concept of a **praxicon** - a motor dictionary of actions. Just as a lexicon stores words and their meanings, a praxicon would store movement patterns and their associated sensory inputs.

When you hear the word 'hammer,' your motor cortex activates - you unconsciously imagine the gripping and swinging motion. This is language grounding: understanding words through the lens of physical action. Current AI lacks this grounding entirely.

The researchers propose building a praxicon for robots - a database of motor representations for every verb in a language. This would allow robots to understand language in a fundamentally human-like way, connecting words to actions rather than just to other words.

How to Build Embodied AI

The paper outlines four approaches to building embodied AI:

1. **Computational simulations** - Creating biomechanically realistic virtual agents in physically realistic virtual worlds

2. **Simplified models** - Building theoretical models that capture the essence of cognitive behavior without needing full biological detail

3. **Animal experiments** - Studying simple organisms (from sea slugs to primates) to understand how embodiment works in nature

4. **Human studies** - Investigating how embodiment shapes human language, cognition, and learning

Is Embodiment the Path to General Intelligence?

Here's a striking fact: the only examples of general intelligence we know of - humans, dolphins, chimpanzees, octopuses - are all embodied. Every intelligent creature we can agree on has a body that interacts with the world.

This suggests that embodiment may be **necessary** (though perhaps not sufficient) for creating truly intelligent agents. Without a body, AI may be limited to what the paper calls simulated general intelligence - impressive on the surface, but fundamentally hollow underneath.

Key Takeaway: Intelligence didn't evolve in a vacuum. It evolved in bodies that interact with physical environments. Giving AI a body may be the missing piece for achieving true understanding.

3. Can AI Really Talk? The Limits of ChatGPT

What LLMs Actually Do

Large language models like ChatGPT are neural networks trained on massive amounts of text. Their goal is simple: predict the next word (or token) in a sequence. When you type a question, the model calculates the probability distribution of possible next words and picks the most likely one.

This approach has produced astonishing results. LLMs can write essays, translate languages, summarize documents, and even write code. But the paper argues that this approach has **fundamental limits** that cannot be overcome by simply making models bigger.

The Biological Plausibility Problem

Current AI models are not biologically, cognitively, or developmentally plausible. They differ from human brains in fundamental ways:

No modularity: The brain has separate systems for language, memory, reasoning, and social cognition. LLMs mash everything into one monolithic network.

No developmental phase: Humans have years of childhood learning through play, social interaction, and exploration. LLMs are trained once on massive data and then barely updated.

No embodied grounding: Human language is rooted in perception and action. LLMs process text in isolation, without any connection to the physical world.

No consciousness: The brain has both conscious (declarative) and unconscious (procedural) learning systems. LLMs have no equivalent of either.

The Fundamental Limits of Language AI

Language serves as a medium for expressing thought, but it is not synonymous with reasoning and thought. When you read a novel, you don't just process the words - you build mental images, feel emotions, and connect the story to your own experiences.

LLMs, by contrast, process words only in relationship to other words. They have no mental images, no emotions, and no personal experiences. This is why they can produce fluent text that is factually wrong, logically inconsistent, or morally problematic.

The Cognitive Atrophy Problem

The paper raises a concerning possibility: **cognitive atrophy from AI reliance**. When we constantly offload thinking to external tools (calculators, GPS, AI assistants), we may be weakening our own cognitive abilities.

There's evidence for this: the 'anti-Flynn effect' - the decline in IQ scores observed in Western countries since the 1970s - appears to correlate with increased cognitive offloading that began with the widespread use of calculators.

If we rely too heavily on AI for thinking, we may lose the very cognitive abilities that make us intelligent in the first place. The paper argues that we need to find ways to use AI as a tool that enhances rather than replaces human cognition.

Can AI Develop Like a Child?

The paper asks a provocative question: should LLMs have a developmental phase similar to early childhood? Humans have a long development during which they learn about the world through embodied interaction, social engagement, and gradual skill building.

What if AI systems had something similar? Instead of being trained once on massive data, they could learn gradually through interaction with the world, building understanding step by step. This could lead to more robust, flexible, and truly intelligent systems.

Key Takeaway: Language is a powerful tool for expressing thought, but it's not the same as thinking. True AI needs to go beyond words - it needs embodied experience, developmental learning, and the ability to understand the world through action.

4. Building Better Robots: Lessons from the Brain

The Efficiency Problem

Deep learning is incredibly powerful at learning patterns from data, but it comes at a cost. Training a large AI model requires enormous amounts of data, energy, and computation. The human brain, by contrast, learns efficiently from relatively few examples and uses a fraction of the energy.

The paper argues that robots need to take inspiration from the brain to learn more efficiently. This means co-designing computation with sensors and actuators, co-locating memory and processing (instead of shuttling data back and forth), and exploiting the physics of the body itself.

The Safety Problem

How can we design robots that are guaranteed to be safe? This is a critical question as robots move from factories into homes, hospitals, and public spaces.

Current deep learning approaches provide flexibility but cannot guarantee safety. Traditional control theory provides safety guarantees but lacks flexibility. The paper argues for a hybrid approach: use safe, predictable control at lower levels and flexible AI at higher levels.

The Three Principles of Brain-Inspired Robotics

The paper identifies three key principles for building better robots:

Principle 1: Hierarchical Stack

The brain has layers of control operating at different time scales - from millisecond reflexes in the spinal cord to long-term planning in the cortex. Robots need a similar hierarchy: fast reflexes at the bottom, flexible planning at the top, and intermediate layers connecting them.

Principle 2: Distributed Control

Mainstream robots use rigid, centralized control. The brain, by contrast, distributes control across many levels. Simple reflexes happen at the spinal cord level without involving the brain. More complex behaviors involve coordination across multiple brain areas. Robots should learn from this distributed architecture.

Principle 3: Full-Stack Codesign

In biology, the body and brain co-evolved. The shape of a bird's wing is inseparable from the neural control that moves it. Robots should be designed the same way - the physical body and the control system should be developed together, not separately.

Why LLMs Aren't Silver Bullets for Robotics

Some researchers hope that large language models will solve robotics. The paper is skeptical. LLMs operate in discrete, turn-based text space. Robotics requires real-time control in a continuous physical world. LLMs are expensive to train, tied to specific hardware, and lack the safety and predictability needed for physical interaction.

Instead, the paper argues for building robots that learn from their own physical experience - developing 'intuitive physics' through interaction with the world, similar to how infants learn.

The Need for Standards

The paper advocates for creating standards for robotic hardware and software - analogous to how USB, Linux, and ARM created ecosystems for computing. An 'operating system' for robotics would allow researchers to build on each other's work rather than reinventing the wheel.

> **Key Takeaway:** The best robots won't be the ones with the biggest neural networks. They'll be the ones that most cleverly combine brain-inspired control with clever body design.

5. How Humans Learn (And How AI Should Too)

Lifelong Learning

Humans learn continuously throughout their lives. We learn to walk as toddlers, read as children, drive as teenagers, and work as adults. Each new skill builds on previous learning without erasing it.

Current AI systems can't do this. When you train a neural network on new data, it typically forgets what it learned before - a problem called 'catastrophic forgetting.' The paper identifies this as one of the most important problems to solve.

The brain uses several mechanisms to enable lifelong learning: **neuromodulation** (chemical signals that gate what gets learned), **varying synaptic timescales** (some connections change quickly for rapid adaptation, others change slowly for stable knowledge), and **sleep-based memory consolidation** (replaying and strengthening important memories while pruning unimportant ones).

Building Robust AI

Current AI is brittle - small perturbations can cause catastrophic failures. A self-driving car might be confused by a sticker on a stop sign. An image classifier might misidentify a panda as a gibbon after adding a tiny amount of noise.

The brain handles perturbations gracefully. If you put on glasses that flip your vision upside down, your brain adapts within days. If you sprain an ankle, your nervous system immediately adjusts your gait to compensate.

The paper suggests building AI with specialized modules inspired by brain circuits: a **predictive control module** (like the cerebellum) for anticipating and correcting disturbances, and a **high-capacity associative memory** (like the hippocampus) for quickly recalling relevant past experiences.

Modular Learning

The brain doesn't learn everything as one monolithic process. Different brain areas specialize in different tasks - visual cortex for seeing, motor cortex for moving, hippocampus for memory. These modules coordinate through loops and pathways.

AI systems could benefit from similar modularity. Instead of training one giant network on everything, we could build systems with specialized modules that handle different aspects of a problem, coordinated by higher-level control systems.

Development and Evolution

Biological intelligence didn't appear overnight. It was shaped by billions of years of evolution and years of individual development. The paper suggests that AI could benefit from similar processes.

Instead of designing a fixed architecture and training it from scratch, we could use principles from developmental biology to let architectures self-organize based on compact sets of rules - much like a genome. This "developmental AI" could bypass the need for exhaustive retraining by starting from structured initializations that already encode useful knowledge.

Common Sense Through Action

How can we build AI with common sense? The paper argues that we need to shift from passive prediction on static data to **active learning through action-conditioned prediction**.

Instead of just predicting what comes next in a text sequence, AI should actively build its world model by observing the consequences of its own behavior. It should try things, see what happens, and update its understanding accordingly - just like a child learning through play.

Energy Efficiency

The brain uses about 20 watts - less than a light bulb. Large AI models use megawatts - enough to power thousands of homes. The paper argues that this efficiency gap can be closed by adopting the brain's core principle: co-locate memory and computation.

Current computers shuttle data back and forth between memory and processor, wasting enormous amounts of energy. By putting memory and processing on the same chip (as the brain does), we could dramatically reduce energy consumption.

> **Key Takeaway:** The brain learns continuously, efficiently, and robustly. AI needs to adopt similar principles: lifelong learning, modular architecture, active exploration, and energy-efficient hardware.

6. Brain Chips: Computers That Think Like Neurons

What Is Neuromorphic Engineering?

Neuromorphic engineering is the field of building computer hardware that mimics the structure and function of biological neural circuits. Instead of using traditional digital computers that process information in sequential steps, neuromorphic chips use artificial neurons and synapses that communicate through electrical spikes - just like the real brain.

The field was pioneered by Carver Mead in 1986, who realized that the analog VLSI (Very Large Scale Integration) technology used to build computer chips could also be used to build artificial neural systems. Since then, neuromorphic engineering has grown into a major research area.

What's Already Neuro-Inspired in AI?

Many elements of current AI are already inspired by neuroscience, even if researchers don't always realize it:

- **Non-linear activations** (ReLU, softmax) - inspired by the way neurons fire
- **Convolutional networks** - inspired by the hierarchical structure of the visual cortex
- **Attention mechanisms** - inspired by how the brain focuses on relevant information
- **Reinforcement learning** - inspired by dopamine reward systems in the basal ganglia
- **Recurrent networks** - inspired by the temporal dynamics of real neural circuits

What's NOT Like the Brain

But there are fundamental differences between current AI and biological brains:

- **Tokenization:** AI breaks input into discrete tokens. The brain processes continuous, streaming signals. Natural perception is smooth and temporal, not chopped into chunks.
- **Feedforward architecture:** Most AI networks flow in one direction. The brain has massive feedback - signals flow both forward and backward through every layer.
- **Dense computation:** AI uses dense matrix multiplication. The brain uses sparse, event-driven computation - neurons only fire when they have something to say.
- **No statefulness:** AI processes each input independently. The brain maintains continuous internal states that evolve over time.

The Role of Spikes

Neurons communicate through electrical spikes - brief pulses of activity. This spiking is fundamentally binary: a neuron either fires or it doesn't. This might seem limiting, but it's actually incredibly efficient. By encoding information in the timing of spikes rather than their amplitude, the brain achieves remarkable energy efficiency.

Neuromorphic chips replicate this spiking behavior. Recent advances have expanded beyond simple binary spikes to 'spike packets' that can carry multi-bit data while preserving the energy efficiency of sparse, asynchronous communication.

The Co-Design Approach

The paper advocates for a co-design approach that integrates four elements: **Neuroscience** (understanding the brain), **Algorithms** (brain-inspired learning rules), **Hardware** (neuromorphic chips), and **Sensors** (event-driven cameras and touch sensors).

Each element informs and improves the others. Better neuroscience reveals new computational principles. Better algorithms demand new hardware capabilities. New hardware enables new sensors. And new sensors generate data that tests neuroscience theories.

The Smartphone NPU Opportunity

An exciting near-term opportunity: the Neural Processing Units (NPUs) in modern smartphones. With over 9 billion smartphones in use worldwide, these NPUs represent the most advanced battery-powered AI processing technology available. They could be repurposed for robotics, IoT devices, and prosthetics.

> **Key Takeaway:** Neuromorphic engineering is building computers that think like brains - using spikes instead of clock cycles, sparsity instead of dense computation, and co-located memory and processing instead of separate chips.

7. The Roadmap: 5, 10, and 20 Years From Now

5-Year Goals (By 2030)

The near-term goals focus on building foundational capabilities and infrastructure:

- **Neuromorphic twin of a mouse connectome** - A complete simulation of a mouse brain's wiring diagram running on neuromorphic hardware
- **Biological realism in AI** - Incorporating more biological neurons, synapses, and neuromodulatory computation into AI systems
- **3D memory integration** - Stacking memory and compute in three dimensions to support sparse, brain-like workloads
- **Open hardware for academia** - Making DRAM interfaces and neuromorphic sensors accessible to university researchers
- **Wearable neuromorphic devices** - Glasses, earables, and health monitors with on-chip neuromorphic processing
- **Smartphone NPUs for robotics** - Opening up smartphone AI chips to robotics and IoT applications
- **Fleet learning** - Robots that learn from each other's experiences, not just their own

10-Year Goals (By 2035)

Medium-term goals build on the foundation with more ambitious targets:

- **Neuromorphic twin of a monkey connectome** - Scaling up to primate brain complexity
- **Bio-compatible neuromorphic interfaces** - Brain-machine interfaces that seamlessly integrate with biological tissue
- **Wearable neuromorphic sensors** - Everyday devices with brain-like processing
- **3D heterogeneous integration** - Combining diverse sensors, memory, and compute in three-dimensional architectures
- **Fully probabilistic AI** - Systems that reason with uncertainty, like the brain
- **Resource-constrained uncertainty quantification** - AI that can assess its own confidence even on small devices

20+ Year Goals (By 2045)

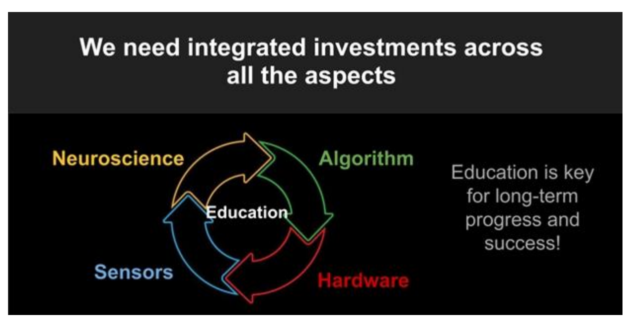
Long-term goals are truly transformative:

- **Neuromorphic twin of a human connectome** - A complete simulation of the human brain's wiring on neuromorphic hardware
- **Closed-loop neural interfaces** - Real-time sensing, computing, and actuation connected to the brain
- **Neuromorphic prosthetics** - Artificial limbs and organs controlled by brain-like chips
- **Sub-kilowatt AI supercomputer** - A system performing at today's cluster levels while consuming less power than a household appliance

Human-level touch and olfaction - Robots that can feel and smell as well as humans

Mass-produced neuromorphic chips - Brain-inspired processors available at consumer prices

The Big Vision: By 2045, we could have neuromorphic computers that think like brains, prosthetics that feel like real limbs, and AI systems that learn as efficiently as a child - all running on chips that use less power than a light bulb.



8. SWOT Analysis: Strengths, Weaknesses, Opportunities, Threats

Strengths

Multidisciplinary community - NeuroAI brings together neuroscience, AI, robotics, and engineering in genuine collaboration

Rich data sources - We now have connectomes, cell types, large-scale electrophysiology, and natural behavior tracking data

Embodiment focus - NeuroAI emphasizes connection to the physical world and computational efficiency

Attracts top talent - The field is exciting enough to draw the best students

Giving back to natural sciences - AI methods are helping neuroscientists make new discoveries

Weaknesses

Lacking theory - Few theorems, proofs, or formal guarantees for NeuroAI approaches

Complexity - Many biological components are extremely challenging to simulate (neuromodulators, peptides, gap junctions)

Inaccessible hardware - Cutting-edge neuromorphic chips are too expensive for most academic labs

Resistance from mainstream AI - Most AI researchers in industry don't care about neuroscience

Training gaps - Most AI researchers are trained in computer science, lacking interdisciplinary insight

Opportunities

New applications - Biomedicine, healthcare, space, edge computing

Next-gen AI paradigbs - Potential to leapfrog current LLMs with biologically-inspired approaches

In vivo - in silico hybrids - Brain-computer interfaces and computational neuroscience platforms

Education reform - New degree programs training students in AI beyond computer science

Policy engagement - Better training for policymakers and scientific representation in government

Threats

Industry dominance - Current AI paradigms may dominate so rapidly that they leave no space for NeuroAI

Talent drain - Without critical mass of young talent, NeuroAI could disappear

Lack of collaboration - Different frameworks may diverge without standards for exchange

Misuse potential - AI could be used for spreading misinformation and privacy/security threats

Funding uncertainty - Long-term, high-risk research is hard to fund in a short-term focused world

9. Voices from the Workshop: What Researchers Really Think

One of the most valuable parts of the paper is the collection of personal statements from the 70 workshop participants. These are the world's leading researchers in neuroscience, AI, robotics, and neuromorphic engineering. Here are some of the most compelling perspectives:

Yann LeCun (NYU / Meta AI)

LeCun, one of the godfathers of deep learning, argues that current AI is fundamentally limited because it doesn't have a world model. He advocates for building AI systems that can learn from observation and interaction, not just from labeled data. His vision: AI that learns the way a child does - by watching, experimenting, and building internal models of how the world works.

Terry Sejnowski (Salk Institute / UCSD)

Sejnowski, a pioneer of computational neuroscience, asks: 'Can large language models think like us?' His answer: not yet, but they're pointing in interesting directions. He emphasizes that LLMs have demonstrated that massive scale can produce emergent capabilities - but true understanding requires grounding in the physical world.

Evelina Fedorenko (MIT)

Fedorenko, who studies the brain's language networks, raises a crucial question: does AI need to be modular like the brain? She argues that separating language from reasoning may be essential for building AI that truly understands, rather than just pattern-matching on text.

Barbara Oakley (Oakland University)

Oakley, an expert on learning and cognition, warns about cognitive atrophy from AI reliance. She presents evidence that our thinking abilities are declining as we offload more cognitive tasks to machines. Her solution: use AI as a tool that enhances learning, not replaces it.

John Doyle (Caltech)

Doyle, a control theory legend, argues that robustness requires layered architectures. He shows how the brain's hierarchical structure - from spinal cord reflexes to cortical planning - provides a blueprint for building robust AI systems that can handle uncertainty and perturbation.

Yiannis Aloimonos (University of Maryland)

Aloimonos proposes a radical idea: a 'praxicon' - a motor dictionary that grounds language in physical action. Just as we understand words through their relationships to other words, robots should understand language through their relationships to physical actions.

Chiara Bartolozzi (Italian Institute of Technology)

Bartolozzi argues that modern AI ignores the most fundamental aspect of intelligence: embodiment. Biological brains are not disembodied entities - they're entangled with bodies that filter, pre-process, and shape information through continuous sensorimotor feedback loops.

Bing Brunton (University of Washington)

Brunton makes a compelling case: the only examples of intelligence we agree on are embodied. Therefore, developing embodied NeuroAI may be necessary for achieving true intelligence. She's building a fully integrated neuromechanical model of the fruit fly to test this idea.

Tobi Delbruck (ETH Zurich)

Delbruck, a pioneer of neuromorphic sensors, observes that 'slavishly imitating the brain was holding us back.' Digital computing continued to scale and became manufacturable at complex scales. But he sees neuromorphic principles increasingly appearing in mass-production hardware - smartphones, wearables, and robots.

Blake Richards (McGill University)

Richards argues that we need AI that learns continuously, like the brain. He advocates for mechanisms like neuromodulation and varied synaptic timescales to enable lifelong learning without catastrophic forgetting.

Gert Cauwenberghs (UCSD)

Cauwenberghs emphasizes the urgency of action: 'The ferocious appetite of today's AI for energy consumption at unsustainable levels endangering our future, compounded by the brittleness of today's AI at even the simplest unmalicious challenges, call for a refocused collective effort in applying neuromorphic principles.'

Jason Eshraghian (UC Santa Cruz)

Eshraghian sees neuromorphic engineers as the bridge between neuroscience and real products. Near-term, they exploit modern silicon with neuromorphic principles to cut computation. Long-term, they need to 'crack the hardware lottery so that the most promising theories of learning and cognition are not buried by the wrong hardware.'

Anthony Zador (Cold Spring Harbor Laboratory)

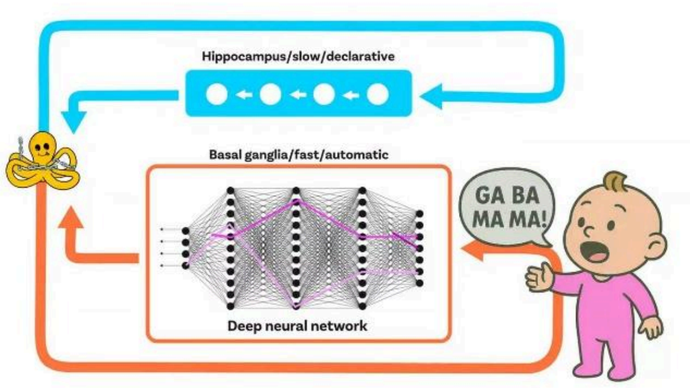
Zador raises a provocative question about what components of biology are actually needed for AI performance. Can we strip away biological complexity and find the essential principles? Or do we need the full richness of biological computation - including glia, neuromodulators, and developmental processes?

Michael Berry (Princeton)

Berry studies how even simple neural circuits like the retina can make predictions about future stimuli. This suggests that predictive computation may be a fundamental feature of all neural circuits - a principle that could transform how we build AI systems.

Chris Kanan (University of Rochester)

Kanan suggests retiring the term 'AGI' and instead identifying two distinct paths for intelligence: one focused on human-like general capabilities, another on narrow but superhuman performance in specific domains. This clarity could help focus research efforts more effectively.



10. Why This Matters to You

For Patients and Families

NeuroAI could revolutionize healthcare. Brain-machine interfaces could restore communication to people with paralysis. Neuromorphic prosthetics could give amputees limbs that feel real. AI-powered diagnostic tools could catch diseases earlier and more accurately.

For Students and Educators

This paper argues for a fundamental change in how we train AI researchers. Instead of computer science silos, we need interdisciplinary programs that combine neuroscience, cognitive science, engineering, and computer science. The next generation of AI breakthroughs will come from people who can think across these boundaries.

For Policymakers

The race for AI dominance is not just about chatbots and search engines. It's about national security, economic competitiveness, and the future of work. Investing in NeuroAI research is investing in the next generation of technology that will shape our society.

For Everyone

The industrial revolution enhanced our physical capabilities. The NeuroAI revolution could substantially enhance our cognitive capabilities while deepening our understanding of our own minds. As the paper concludes: 'We are at the start of a new era. Our field is fundamentally changing. Things that were literally impossible a couple of years ago are now becoming routine. And we are only beginning to see what AI and quantum computing will make possible.'

The Bottom Line: The brain is the most remarkable computing system in the known universe. By understanding how it works - and building AI that respects those principles - we can create technology that is more efficient, more robust, and more truly intelligent than anything we have today.

11. Key Terms Explained

NeuroAI

A field that combines neuroscience and AI to build better artificial intelligence by learning from how the brain works.

Embodiment

The idea that intelligence requires a body that interacts with the physical world, not just software processing text and images.

Neuromorphic Engineering

Building computer hardware that mimics the structure and function of biological neural circuits, using artificial neurons and synapses.

LLM (Large Language Model)

AI systems like ChatGPT trained on massive text data to predict the next word in a sequence.

Connectome

A complete map of the neural connections in a brain - the 'wiring diagram' of the nervous system.

Catastrophic Forgetting

The tendency of neural networks to forget previously learned information when trained on new data.

Neuromodulation

Chemical signals in the brain (like dopamine) that regulate which connections get strengthened or weakened during learning.

Praxicon

A hypothetical 'motor dictionary' that stores movement patterns and their associated sensory inputs - the action equivalent of a word dictionary.

Spiking Neural Networks

Neural networks that communicate through discrete electrical spikes, like biological neurons, rather than continuous values.

Hierarchical Control

A system with multiple layers of control operating at different time scales - from fast reflexes to slow planning.

Von Neumann Architecture

The standard computer design where memory and processing are separate, requiring data to be shuttled back and forth.

LLM (Large Reasoning Model)

An advanced LLM that can perform multi-step reasoning, not just next-word prediction.

AGI (Artificial General Intelligence)

Hypothetical AI with human-level intelligence across all cognitive tasks.

Neural Processing Unit (NPU)

Specialized AI chips built into modern smartphones for efficient on-device AI processing.

Flynn Effect

The observed rise in IQ scores over the 20th century, which has recently reversed (the 'anti-Flynn effect').

12. Deep Dive: How the Brain Actually Works

Neurons and Synapses: The Building Blocks

The human brain contains approximately 86 billion neurons, each connected to thousands of others through synapses. A neuron receives electrical signals from other neurons through its dendrites, processes them in its cell body, and if the combined input exceeds a threshold, it fires an electrical spike down its axon to communicate with the next neurons in the chain.

This is fundamentally different from how computers work. Computers process information in clock cycles - synchronized ticks that coordinate all operations. Neurons operate asynchronously - they fire when they have something to say, not when a clock tells them to. This makes the brain incredibly energy-efficient: instead of processing every possible calculation, neurons only compute when there's actually something to compute.

Neuromodulation: The Brain's Chemical Messages

Beyond electrical signals, the brain uses chemical messages called neuromodulators to coordinate activity across large brain areas. The most famous is dopamine, which signals reward and helps the brain decide what's worth learning. But there are many others:

- **Serotonin:** Regulates mood, sleep, and social behavior
- **Norepinephrine:** Controls attention and arousal - helps you focus during important tasks
- **Acetylcholine:** Enhances learning and memory formation
- **GABA:** The brain's main inhibitory signal - prevents over-excitation
- **Glutamate:** The brain's main excitatory signal - enables communication between neurons

These chemical systems don't just transmit information - they gate what gets learned and when. When dopamine signals that something is important, the brain strengthens the connections involved. When norepinephrine signals novelty, the brain pays closer attention. Current AI has no equivalent of this chemical gating - one of the key things missing from artificial neural networks.

The Hierarchical Brain: From Reflexes to Planning

The brain is organized in layers, from the most basic reflexes at the bottom to the most sophisticated cognition at the top. Understanding this hierarchy is essential for building better AI and robots.

Spinal cord level: The simplest behaviors happen here. Touch a hot stove and your hand pulls back before your brain even knows what happened. These spinal reflexes operate in milliseconds and don't require any conscious thought.

Brainstem level: Breathing, heart rate, basic movement coordination. These are automatic but can be modulated by higher brain areas (you can hold your breath, for example).

Cerebellum level: Motor learning and coordination. The cerebellum is like a precision tuner - it detects errors in movement and gradually adjusts the controllers to make movements smoother and more accurate.

Basal ganglia level: Habit formation and action selection. This is where procedural memory lives - the kind of knowledge that lets you ride a bike without thinking about it.

Cortex level: Conscious thought, planning, language, abstract reasoning. This is the most recently evolved part of the brain and the most different between humans and other animals.

Current AI systems typically implement only one level of this hierarchy. The paper argues that robust AI needs all levels working together - fast reflexes at the bottom, flexible planning at the top, and coordination between them.

Sleep and Memory Consolidation

One of the brain's most remarkable abilities is memory consolidation during sleep. While you sleep, your hippocampus replays the day's experiences, strengthening important memories and pruning unimportant ones. This process happens primarily during slow-wave sleep and REM sleep.

Current AI has no equivalent of sleep. When you train a neural network, it processes all data equally - there's no mechanism for deciding what's important enough to keep long-term. The paper suggests that incorporating sleep-like consolidation mechanisms could help AI systems learn more efficiently and avoid catastrophic forgetting.

13. Deep Dive: Specific Research Projects

The Fruit Fly Connectome Project

One of the most exciting projects mentioned in the paper is the complete mapping of the fruit fly brain. In 2024, scientists completed the first full connectome of an adult insect brain - every neuron and every synapse mapped in detail.

The fruit fly is a powerful model organism because its brain is small enough to map completely (about 100,000 neurons and 10 million synapses) but complex enough to exhibit sophisticated behaviors like learning, memory, navigation, and social interaction.

Researchers are now using this connectome to build computational models that simulate the fly's neural circuits. By connecting these neural models to biomechanical models of the fly's body in physics simulators, they can test whether the neural circuits produce realistic behavior.

This approach - building a complete brain-body model from the ground up - could eventually be scaled up to larger organisms and eventually to humans. It represents a fundamentally different approach to AI: instead of hand-designing neural networks, we could grow them from biological blueprints.

The Telluride Workshop: 30 Years of Neuromorphic Innovation

The paper mentions the Telluride Neuromorphic Cognition Engineering Workshop, which has been running for over 30 years. This annual summer workshop brings together researchers from neuroscience, computer science, and electrical engineering to develop new neuromorphic computing technologies.

At Telluride, participants build working neuromorphic systems in just a few weeks. These workshops have produced many of the key innovations in the field, including new types of silicon retinas, cochleas, and neural processors. The workshop has trained generations of researchers who now lead both academic and industry neuromorphic programs.

The BrainGate2 Communication Neuroprosthesis

While not directly discussed in the workshop, the paper's vision of brain-machine interfaces connects to real clinical work like BrainGate2. This system uses tiny electrode arrays implanted in the motor cortex to decode neural signals into cursor movements on a computer screen.

Recent advances have enabled paralyzed patients to type at speeds approaching normal handwriting by decoding imagined handwriting movements. This is exactly the kind of practical application that NeuroAI research aims to enable.

Neuromorphic Chips: Loihi, TrueNorth, and Beyond

Several neuromorphic chips are already being developed and tested:

Intel Loihi 2: Uses asynchronous spiking neurons with on-chip learning. Can simulate millions of neurons with biological-like plasticity. Used for optimization, pattern recognition, and robotics.

IBM TrueNorth: 1 million digital neurons and 256 million synapses on a chip the size of a postage stamp, using only 70 milliwatts of power.

SpiNNaker 2: The University of Manchester's massive neural simulation platform, designed to simulate up to a billion neurons in real time.

BrainScaleS: An analog neuromorphic system that operates 1,000 times faster than real-time, allowing researchers to study neural dynamics at accelerated timescales.

These chips are still far from matching the brain's capabilities, but they demonstrate the feasibility of building hardware that computes like a neural system rather than like a traditional computer.

The Swiss/US Neuromorphic Computing Initiative

The paper notes that China has made substantial investments in AI, robotics, and neuroscience. In response, the authors call for renewed US investment in neuromorphic AI engineering, including continued support for programs like the Telluride Workshop and new initiatives in neuromorphic hardware development.

This is not just an academic concern - neuromorphic computing has direct applications in national security, healthcare, autonomous systems, and economic competitiveness. Countries that lead in this technology will have significant advantages in the coming decades.

14. Deep Dive: Comparing AI Approaches

Current AI vs. NeuroAI: A Side-by-Side Comparison

To understand why NeuroAI matters, it helps to compare the current dominant approach to AI with the NeuroAI vision:

Dimension	Current AI (Deep Learning)	NeuroAI Vision
Architecture	Monolithic, feedforward	Hierarchical, distributed, with feedback
Learning	Batch training on fixed datasets	Continuous lifelong learning
Energy	Megawatts (data center scale)	Watts (brain scale)
Data	Billions of labeled examples	Few examples, active exploration
Robustness	Brittle to perturbations	Graceful degradation and adaptation
Memory	No long-term memory	Sleep-based consolidation
Body	Software only	Embodied in physical world
Development	No developmental phase	Child-like developmental learning
Hardware	GPUs (dense digital)	Neuromorphic (sparse, analog)

Why Scaling Alone Won't Work

A common argument in AI is that bigger models are better models. If we just make neural networks large enough and train them on enough data, they'll eventually achieve human-level intelligence.

The paper challenges this assumption. It points out that current AI models are fundamentally limited in ways that can't be fixed by scaling:

- **No embodiment:** Adding more parameters won't give AI a body or physical experience
- **No development:** Training longer won't simulate years of childhood learning
- **No energy efficiency:** Making models bigger only makes them more power-hungry
- **No feedback:** Adding more layers won't create the massive feedback loops of the brain
- **No robustness:** More data won't make AI robust to novel perturbations

Instead, the paper argues for fundamentally new approaches inspired by neuroscience - not just bigger versions of the same approaches we already have.

The Two Paths Forward

The paper identifies two complementary paths for advancing AI:

Path 1: Improve current AI with neuroscience insights. Take specific principles from neuroscience - neuromodulation, hierarchical control, feedback loops - and incorporate them into existing AI architectures. This is incremental progress that can happen relatively quickly.

Path 2: Build new AI from biological principles. Start from scratch, using neuroscience as a blueprint. Build neuromorphic hardware, developmental learning algorithms, and embodied agents that learn through physical interaction. This is revolutionary progress that will take longer but could lead to fundamentally better systems.

The paper advocates for pursuing both paths simultaneously. Path 1 delivers near-term improvements while Path 2 develops the breakthroughs needed for long-term progress.

15. Deep Dive: The Ethics of NeuroAI

Privacy and Neural Data

As brain-machine interfaces become more sophisticated, they'll be able to read increasingly detailed neural information. This raises serious privacy concerns: who owns your neural data? Can employers access it? Can it be used against you in court?

The paper calls for developing ethical and legal standards before these technologies become widespread. Current privacy laws were not designed for continuous neural monitoring, and new frameworks are urgently needed.

Autonomy and Control

If we build AI systems that think more like brains - with emotions, motivations, and self-directed learning - questions of autonomy become more pressing. How do we ensure that such systems remain aligned with human values?

The paper suggests applying insights from neuroscience about cognitive biases and motivated reasoning to develop better alignment strategies. Understanding how human brains can exhibit pathological altruism (well-intended actions that cause harm) can inform AI safety approaches that avoid similar failure modes.

The Dual-Use Dilemma

NeuroAI technologies have enormous potential for good - restoring communication to paralyzed patients, treating neurological disorders, advancing scientific understanding. But they also have potential for misuse - surveillance, manipulation, or weaponization.

The paper notes that AI, both neuro-inspired and deep learning-based, could be misused for spreading wrong information and noise, and for privacy and security threats both digital and physical. Ethical and legal considerations should be developed to ensure NeuroAI is not inadvertently or intentionally misused.

Accessibility and Equity

Advanced neurotechnology is currently concentrated in wealthy countries and elite research institutions. If NeuroAI benefits are to be shared broadly, deliberate efforts must be made to ensure accessibility.

The paper advocates for open-source neuromorphic hardware, affordable brain-machine interfaces, and educational programs that train researchers from diverse backgrounds. The goal is a NeuroAI revolution that benefits all of humanity, not just a privileged few.

16. Deep Dive: Practical Applications

Healthcare and Medicine

NeuroAI could transform healthcare in several ways:

- **Brain-machine interfaces:** Restoring communication and mobility to people with paralysis, locked-in syndrome, or severe motor disabilities
- **Neuromorphic prosthetics:** Artificial limbs that feel natural and can be controlled by thought, with sensory feedback that allows the user to feel pressure and temperature
- **Neurological disorder treatment:** Closed-loop brain stimulation systems that detect and treat seizures, tremors, or depressive episodes in real time
- **AI diagnostics:** Medical imaging analysis that combines the pattern recognition of deep learning with the robustness and interpretability of brain-inspired algorithms

Robotics and Autonomous Systems

NeuroAI principles could create robots that are:

- **More energy-efficient:** Using neuromorphic chips and event-driven computation to operate for hours or days on a single battery charge
- **More robust:** Handling unexpected situations gracefully instead of failing catastrophically
- **More adaptable:** Learning new tasks quickly from few examples, like a human
- **More natural:** Interacting with people in ways that feel intuitive rather than mechanical

Education and Learning

NeuroAI insights could revolutionize education by:

- **Personalized learning:** AI tutors that adapt to each student's learning style and pace, drawing on neuroscience research about how people actually learn
- **Cognitive enhancement:** Tools that strengthen rather than replace cognitive abilities, helping students develop critical thinking and problem-solving skills
- **Teacher training:** Using NeuroAI to help teachers understand how their students' brains process information

Environmental Monitoring

Neuromorphic sensors and processors could enable:

- **Ultra-low-power environmental monitoring:** Networked sensors that run for years on battery power, detecting changes in ecosystems
- **Wildlife tracking:** Miniature neuromorphic processors that can identify and track individual animals from audio or video
- **Climate modeling:** More efficient AI algorithms for simulating complex climate systems

17. The Complete SWOT Analysis

From the Researchers' Perspective

The workshop participants conducted a thorough SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis of the NeuroAI field. Here's a more detailed look at each category:

Strengths (Detailed)

Collaborative community: Workshops, interdisciplinary grants, and co-mentorship create a supportive research environment

Multiple funding sources: NSF, NIH, DoD, foundations, and industry all support NeuroAI

Rich neuroscience data: Connectomes, cell types, large-scale electrophysiology, and natural behavior tracking provide unprecedented research material

Embodiment focus: Connection to the physical world and computational efficiency address key limitations of current AI

Talent magnet: The previous generation of NeuroAI (today's AI) has demonstrated its potential, attracting top students

Giving back: AI methods are helping neuroscientists make new discoveries in their fields

Weaknesses (Detailed)

No theoretical foundation: Few theorems, proofs, or formal guarantees for NeuroAI approaches

Biological complexity: Neuromodulators, peptides, gap junctions are extremely challenging to simulate or mimic

Jargon overload: The field uses inconsistent terminology, making it hard for newcomers

Hardware access: Cutting-edge neuromorphic chips are too expensive for most academics

Industry resistance: Most AI researchers in industry don't prioritize neuroscience insights

Training gaps: Computer science programs weren't designed for modern AI, let alone NeuroAI

Opportunities (Detailed)

New applications: Biomedicine, healthcare, space exploration, edge computing

Next-gen paradigms: Potential to leapfrog current LLMs with biologically-inspired approaches

In vivo-in silico hybrids: Brain-computer interfaces and computational neuroscience platforms

Education reform: New degree programs training students beyond computer science silos

Policy engagement: Better training for policymakers and scientific representation in government

Hypothesis generation: NeuroAI will suggest new hypotheses about brain organization that can be tested experimentally

Threats (Detailed)

Industry dominance: Current AI paradigms may leave no space for NeuroAI alternatives

Talent drain: Without critical mass, NeuroAI could disappear as a field

- **Divergent frameworks:** Different groups may develop incompatible approaches without standards
- **Misuse potential:** AI could spread misinformation or threaten privacy and security
- **Funding uncertainty:** Long-term, high-risk research is hard to fund
- **Regulatory lag:** Laws predate high-resolution, continuous neural signal acquisition

From the Students' and Postdocs' Perspective

The workshop also gathered input from the next generation of NeuroAI researchers. Their perspective is crucial because they'll be the ones to carry the field forward.

Student Strengths

- Network-style grants enabling multi-institution collaboration
- Training programs like Telluride, Neuromatch, and NeuroPAC
- Interdisciplinary research bridging theory and experiment
- Student fellowships that provide independence

Student Weaknesses

- Siloed academic departments making cross-department collaboration difficult
- Insufficient ethics/safety education
- Limited public outreach and feedback mechanisms

Student Opportunities

- Funding for lab exchanges and international collaboration
- Competitions that incentivize improvement and collaboration
- Shared compute resources from industry to academia
- Revival of MOSIS-style chip design programs

Student Threats

- Funding uncertainty and visa/immigration issues
- Brain drain to industry and international competitors
- Postdoc compensation that doesn't reflect the value of the work

18. The Workshop in Detail

How the Workshop Was Organized

The NSF workshop was held on August 27-29, 2025 in Arlington, Virginia. It brought together 70 researchers from five main working groups, each focused on a different aspect of NeuroAI:

- **Embodied Cognition and Computation:** Led by Hillel Chiel, Jean-Marc Fellous, Steve Zucker, Joe Paton, and Alexander Ororbia
- **Language and Communication:** Led by Barbara Oakley, Xiaoqin Wang, Evelina Fedorenko, Cornelia Fermuller, and Shihab Shamma
- **Robotics:** Led by Chiara Bartolozzi, Carmen Amo Alonso, Yulia Sandamirskaya, and John Doyle
- **Learning in Humans and Machines:** Led by Blake Richards, Tony Zador, Terrence Sejnowski, Yann LeCun, Michael Berry, and Mike Stryker
- **Neuromorphic AI Engineering:** Led by Tobi Delbruck, Brad Aimone, Ralph Etienne-Cummings, Gert Cauwenberghs, Abhronil Sengupta, Jason Eshraghian, and Gina Adam

The Three-Day Format

The workshop followed an intensive three-day format:

Day 1 (Wednesday): Opening reception, keynote by Carmen Amo Alonso on 'NeuroAI Meets Control Theory,' and initial position statements from each working group.

Day 2 (Thursday): Morning breakout sessions where working groups discussed their topics in depth. Afternoon interim position statements and training/workforce development discussion. Evening working dinner.

Day 3 (Friday): Second breakout session, final position statements, closing discussion, and farewell.

The Working Group Structure

Each working group was designed to include researchers from multiple disciplines. For example, the Embodiment group included biologists studying sea slugs, roboticists building fruit fly models, and computer scientists developing new algorithms. This cross-pollination was intentional — the organizers believed that breakthroughs would come from unexpected connections between fields.

The workshop also included scribes from each group who documented the discussions, and the entire event was designed to produce not just academic papers but actionable research roadmaps.

19. Personal Statements: Extended Perspectives

Ralph Etienne-Cummings (Johns Hopkins)

Etienne-Cummings sees NeuroAI and neuromorphic engineering as converging into a transformative frontier. He highlights the potential for diagnosing and treating neurological, psychiatric, and neuromuscular conditions through adaptive, implantable, and wearable systems.

His key concern: academia's struggle to compete with industry for talent and resources. Without deliberate effort to support academic NeuroAI research, the field risks being dominated by industry priorities that may not align with long-term scientific progress.

Gina Adam (George Washington University)

Adam emphasizes the need for new neuromorphic hardware technologies that can interface with sensors and actuators efficiently. She's concerned about the lack of research infrastructure for interdisciplinary neuromorphic work — the hardware is too expensive for most labs.

Her vision: a community effort to define initial research goals and build the infrastructure needed for neuromorphic integration and embodied AI.

Brad Aimone (Sandia National Labs)

Aimone declares that 'the days of weak brain-inspiration are over.' With exascale supercomputers and full-scale connectomes now available, we can finally test and falsify bad ideas about how the brain works.

He advocates for algorithms that embrace — not avoid — the brain's heterogeneity, parallelism, stochasticity, and spatio-temporal complexity. True brain-inspired algorithms are 'right around the corner.'

Akwasi Akwaboah (Johns Hopkins)

Akwaboah proposes a 'coupled hierarchical space-time perspective' for modeling intelligent systems. He argues that the continuous dynamics of diffusion and wave propagation, recently discovered in cortical dynamics, may be the underlying modality of perception and cognition.

He's critical of modern AI's 'obsession with tokenization,' arguing that it fails to capture the temporal associations that are fundamental to biological intelligence.

Carmen Amo Alonso (Stanford)

Alonso, trained in aerospace engineering and control theory, brings a unique perspective. She argues that the rise of deep learning has overshadowed the equally impressive progress in control theory — which provides the reliability, scalability, and safety guarantees that AI needs.

Her key insight: understanding a system's dynamics enables principled design of learning systems, leading to more robust and efficient designs. She advocates for a full-stack approach using systems architecture theory.

Chiara Bartolozzi (Italian Institute of Technology)

Bartolozzi argues that foundational models — the most successful AI systems — ignore the most fundamental aspect of intelligence: embodiment. Biological brains are not disembodied entities; they're entangled with bodies that filter and shape information.

She envisions a new perspective on intelligence as embodied and action-directed, which could steer research toward solutions that work more efficiently, using less computational resources and less data.

John Doyle (Caltech)

Doyle, a legendary control theorist, introduces the concept of Universal Layered Architectures (ULA). He argues that extreme robustness requirements in networked systems require layered control architectures that are ubiquitous but poorly understood.

His key contribution: System Level Synthesis (SLS), a new theory that makes lower-layer feedback control with delays, sparsity, and locality both scalable and robust. This could revolutionize how we design both AI systems and robots.

16. Understanding Neural Plasticity

Why the Brain Learns Differently Than AI

One of the most profound differences between the brain and artificial neural networks is how they learn. In current AI, the architecture is fixed before training begins — the network has a predetermined number of neurons and connections, and learning only adjusts the strength of these connections. The brain works completely differently.

Biological neural networks are constantly rewiring themselves in response to experience. This process, called structural plasticity, involves growing new synaptic connections, pruning unused ones, and even creating entirely new neurons in certain brain regions. The brain's architecture is not fixed — it's a living, evolving structure.

This plasticity operates at multiple time scales. On the fastest timescale (milliseconds), synapses can change their strength through short-term facilitation and depression. On medium timescales (hours to days), structural changes occur as new connections form and old ones weaken. On the longest timescales (months to years), entire brain regions can reorganize in response to sustained experience.

Synaptic Plasticity Rules

The basic rules of synaptic plasticity were discovered by Donald Hebb in 1949: neurons that fire together wire together. But modern neuroscience has revealed much more nuance:

> **Spike-Timing Dependent Plasticity (STDP):** The precise timing of spikes determines whether a synapse strengthens or weakens. If the pre-synaptic neuron fires just before the post-synaptic neuron, the connection strengthens (Hebbian learning). If the order is reversed, the connection weakens (anti-Hebbian learning).
>
> **Homeostatic Plasticity:** When a neuron becomes too active, it reduces its sensitivity. When it becomes too quiet, it increases sensitivity. This keeps the brain stable despite constant plasticity.
>
> **Synaptic Scaling:** The brain can globally adjust the strength of all connections to a neuron, maintaining a target firing rate. This is like an automatic volume control for neural circuits.
>
> **Spike-Tagging:** When a synapse is recently active, it gets "tagged" as eligible for long-term strengthening. This ensures that only relevant synapses are modified during learning.

Implications for AI

Current AI doesn't have any of these plasticity mechanisms. Learning in neural networks is achieved through a single rule — backpropagation — that adjusts connection strengths by propagating error signals backward through the network. While this is mathematically elegant, it's not how the brain learns at all.

The paper argues that incorporating multiple plasticity rules — Hebbian learning, homeostatic regulation, and synaptic tagging — could make AI more efficient, robust, and capable of continual learning. Some researchers are already exploring these ideas in 'local learning rules' that don't require the global coordination of backpropagation.

17. Cross-Species Intelligence

What Animal Brains Teach Us

One of the most powerful arguments for NeuroAI is the diversity of biological intelligence. Nature has evolved dozens of fundamentally different solutions to the problem of intelligence, and studying these different solutions reveals principles that might be missed by studying only the human brain.

Insect Intelligence: Small but Mighty

Insects have tiny brains — a honeybee has fewer than a million neurons, compared to the human brain's 86 billion. Yet they demonstrate remarkable cognitive abilities:

- **Navigation:** Bees can navigate complex 3D environments, remember locations of flowers, and optimize their routes — all with a brain smaller than a sesame seed
- **Learning:** Bumblebees can learn to pull strings, solve puzzles, and even learn by observing other bees do these tasks
- **Communication:** The honeybee waggle dance is a symbolic communication system that encodes the direction, distance, and quality of food sources
- **Social cognition:** Ants collectively solve optimization problems that would challenge powerful computers

The paper emphasizes that insect intelligence is not just 'cute' — it's a goldmine of computational principles. The fruit fly connectome project (mentioned earlier) is the first step in reverse-engineering these principles.

Bird Intelligence: A Different Architecture

Birds present a fascinating puzzle. Their brains are organized very differently from mammalian brains — they lack the six-layered cortex that characterizes the human brain. Yet corvids (crows, ravens, jays) demonstrate cognitive abilities that rival those of great apes.

Corvids can plan for the future, use tools, deceive others, and even understand cause-and-effect relationships. The key insight is that their brains achieve these feats with a completely different architecture — a 'nuclear pallium' instead of a layered cortex.

This suggests that intelligence doesn't require the specific architecture of the human brain. Different neural architectures can achieve similar cognitive abilities through different computational principles. For NeuroAI, this means there may be multiple paths to building intelligent machines.

Octopus Intelligence: The Alien Among Us

The octopus represents perhaps the most radical alternative to human-like intelligence. With about 500 million neurons distributed across its body (two-thirds are in its arms), the octopus has a fundamentally different computational architecture.

- **Distributed processing:** Each arm can sense, decide, and act independently, without input from the central brain
- **Self-recognition:** Octopuses can recognize themselves in mirrors, suggesting self-awareness
- **Problem solving:** They can open jars, navigate mazes, and use coconut shells as portable shelters
- **Camouflage:** They can change their skin color and texture in milliseconds, demonstrating real-time visual processing and motor control

Octopus intelligence suggests that consciousness and self-awareness may be possible with very different neural architectures than the human brain. This is both scientifically fascinating and practically relevant — if there are multiple ways to build intelligence, there may be multiple paths to building AI.

18. The Neuroscience of Decision-Making

How the Brain Makes Choices

Understanding how the brain makes decisions is critical for building AI that can navigate the real world. The paper discusses several key insights from neuroscience research on decision-making:

Value-based decisions: When choosing between options, the brain assigns a 'value' to each option based on past experience, current needs, and future goals. These values are computed by neural circuits in the ventromedial prefrontal cortex and ventral striatum.

Evidence accumulation: For perceptual decisions (is this a cat or a dog?), neurons in the lateral intraparietal cortex gradually accumulate evidence until a threshold is reached. This 'drift-diffusion' process explains why decisions take time and become more accurate with longer deliberation.

Explore-exploit tradeoff: The brain must balance exploring new options (which might be better) with exploiting known options (which are guaranteed). Dopamine signals play a key role in regulating this balance.

Contextual Influence on Decisions

One of the most surprising findings in neuroscience is how dramatically context influences decisions. The same option can be valued differently depending on what other options are available — a phenomenon called 'context-dependent valuation.'

For example, people are willing to pay more for a wine when told it comes from an expensive bottle, even if the wine is identical. The brain's value computation is not absolute — it's relative to the current context.

Current AI systems are typically context-independent — they process each input in isolation. The paper suggests that incorporating context-dependent processing could make AI more flexible and human-like in its decision-making.

Emotional Influence on Decisions

The famous case of Phineas Gage — a railroad worker who survived an iron rod through his head — demonstrated the critical role of emotion in decision-making. After his injury destroyed his orbitofrontal cortex, Gage's personality changed dramatically: he became impulsive, made poor decisions, and lost his social skills.

Modern neuroscience has confirmed that emotion is not the enemy of rationality — it's essential for it. Antonio Damasio's 'somatic marker hypothesis' argues that emotional signals guide decision-making by providing a rapid, intuitive assessment of options.

Current AI lacks this emotional guidance system. It can optimize any mathematical objective, but it has no way to evaluate whether that objective is worth pursuing. Adding artificial emotional signals — mechanisms that bias AI toward decisions that are 'good' in a holistic sense — could help align AI with human values.

19. Computational Models of the Brain

Types of Brain Models

The paper discusses several types of computational models that are used to understand the brain and inspire new AI approaches:

Biophysical models: These simulate the detailed electrical behavior of individual neurons, including ion channels, synaptic transmission, and dendritic integration. They're the most realistic but also the most computationally expensive.

Spiking neural network models: These simulate networks of neurons that communicate through discrete spikes, capturing the temporal dynamics of biological neural circuits without modeling every biophysical detail.

Rate-coded models: These simplify neural activity to average firing rates, ignoring the timing of individual spikes. They're less realistic but more computationally tractable.

Abstract models: These capture high-level principles of brain function (like reinforcement learning or predictive coding) without directly simulating neurons at all.

The Tradeoff Between Realism and Tractability

One of the key challenges in NeuroAI is finding the right level of abstraction. Too detailed, and the models become computationally intractable. Too abstract, and they lose the specific neural mechanisms that make biological intelligence work.

The paper advocates for a 'middle ground' approach: models that capture the essential features of neural computation (temporal dynamics, plasticity, feedback, neuromodulation) without getting bogged down in biophysical detail. The fruit fly connectome project is an example of this approach — detailed enough to be biologically realistic, but small enough to be computationally tractable.

Validation and Falsification

A crucial point made by Brad Aimone is that computational models must be falsifiable — they must make predictions that can be tested against experimental data. A model that fits existing data but makes no new predictions is not science; it's curve-fitting.

The paper calls for tighter collaboration between computational modelers and experimental neuroscientists. Models should suggest experiments, and experimental results should constrain models. This cycle of prediction and validation is how science advances.

20. Detailed Research Roadmap

Year 1-2: Foundation Building

The first phase focuses on building the infrastructure needed for NeuroAI research:

Training programs: Establish interdisciplinary PhD programs that combine neuroscience, computer science, and engineering. Create postdoctoral fellowships that support researchers working across disciplines.

Open hardware: Develop affordable neuromorphic chips that academic labs can use. Create standardized interfaces for connecting neuromorphic hardware to sensors and actuators.

Data sharing: Build repositories for sharing connectomes, neural recordings, and behavioral data. Establish data formats and APIs that enable cross-lab collaboration.

Standards: Define community standards for neural simulation frameworks, model evaluation metrics, and benchmarking protocols.

Year 3-5: Breakthrough Research

The second phase focuses on achieving scientific breakthroughs:

Complete connectomes: Complete connectomes for model organisms that demonstrate complex behavior (zebrafish, mouse, primate cortex). Use these to build functional models that can predict behavior.

Neuromorphic systems: Build neuromorphic systems that match biological performance on specific tasks (pattern recognition, motor control, navigation).

Embodied AI: Create robots that learn continuously from interaction with the world, using developmental learning algorithms inspired by child development.

Brain-machine interfaces: Develop BMI systems that enable high-bandwidth communication between brains and computers, with applications in healthcare and research.

Year 5-10: Integration and Application

The third phase focuses on integrating breakthroughs into practical applications:

Neuromorphic computing platforms: Commercial neuromorphic chips that can be used in consumer devices, autonomous vehicles, and industrial applications.

Neuro-inspired AI algorithms: New algorithms that combine the efficiency of neuromorphic computing with the expressiveness of deep learning.

Clinical applications: Brain-machine interfaces, neuroprosthetics, and closed-loop stimulation systems that help patients with neurological conditions.

Educational tools: AI tutors that adapt to individual students using principles from neuroscience research on learning and memory.

Year 10+: The NeuroAI Revolution

The long-term vision is a world where AI systems are as efficient, robust, and adaptable as biological brains:

- **Autonomous agents:** Robots that can learn any new task quickly from few examples, adapting to novel environments without retraining.
- **Brain-computer symbiosis:** Technology that enhances human cognition rather than replacing it, creating a true partnership between human and artificial intelligence.
- **Sustainable computing:** Neuromorphic hardware that uses a fraction of the energy of current AI systems, making advanced AI accessible to everyone.
- **Scientific discovery:** NeuroAI systems that can form hypotheses, design experiments, and discover new principles in neuroscience, creating a virtuous cycle of discovery.

21. Final Thoughts: The Path Ahead

What Needs to Happen

Based on the workshop discussions, the paper identifies several concrete actions needed to advance NeuroAI:

1. **Create interdisciplinary training programs** that combine neuroscience, cognitive science, engineering, and computer science — with student exchanges between labs
2. **Develop open neuromorphic hardware platforms** accessible to academic researchers, similar to how GPUs became accessible for AI research
3. **Establish community standards** for neural simulation frameworks, data formats, and model architectures to enable collaboration
4. **Increase federal funding** for long-term, high-risk NeuroAI research through NSF, NIH, and DARPA
5. **Build industry-academia partnerships** that give academics access to compute resources while giving industry access to talent
6. **Develop ethical frameworks** for neural data privacy, brain-machine interface regulation, and AI safety
7. **Create public engagement programs** to build understanding and support for NeuroAI research

The Timeline

The paper acknowledges that transformative progress will take time. The 5-year goals are achievable with current technology and funding. The 10-year goals require sustained investment and international collaboration. The 20+ year goals are truly visionary — they depend on breakthroughs we can't yet predict.

But the authors are optimistic. As Terry Sejnowski writes: 'The science and engineering of learning is still in its infancy, and the bridge between AI and brains is accelerating advances. Just as the industrial revolution greatly enhanced our physical power, the NeuroAI revolution could significantly enhance our cognitive power and give us a better understanding of ourselves.'

A Call to Action

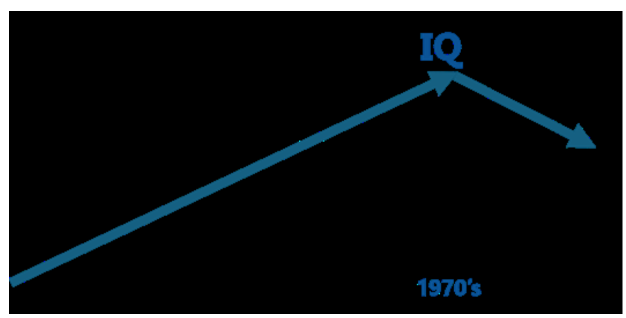
This paper is not just an academic report — it's a call to action. The authors are urging the scientific community, funding agencies, and policymakers to recognize the transformative potential of NeuroAI and invest in it accordingly.

The window of opportunity may be narrow. As John Doyle notes, current AI paradigms may dominate so rapidly that they leave no space for alternatives. If NeuroAI doesn't establish itself soon, it could be crowded out by industry-driven approaches that optimize for short-term profits rather than long-term understanding.

The choice is ours: do we want AI that merely processes text, or AI that truly understands the world? Do we want robots that are brittle and dangerous, or robots that are robust and helpful? Do we want technology that drains our cognitive abilities, or technology that enhances them?

The NeuroAI revolution won't happen automatically. It requires deliberate investment, interdisciplinary collaboration, and a willingness to challenge the dominant paradigm. But if we rise to the challenge, the rewards could be extraordinary.

The Bottom Line (Final): The brain is nature's proof that intelligent, efficient, robust computation is possible. NeuroAI is the path to building technology that matches — and eventually surpasses — what evolution has achieved. The journey will be long, but the destination is worth it.



21. NeuroAI vs. Traditional AI: A Deeper Comparison

The Efficiency Gap

Perhaps the most striking difference between biological and artificial intelligence is energy efficiency. The human brain operates on about 20 watts of power — the same as a dim lightbulb — yet it can recognize faces, understand language, navigate complex environments, and perform countless other cognitive tasks.

By contrast, training a single large language model can consume megawatts of power, equivalent to the energy use of hundreds of homes. Running inference on these models requires additional energy with every query. The gap between biological and artificial efficiency is not just impressive — it's alarming.

This efficiency gap has practical consequences. As AI models get larger, they become more expensive to run, more polluting, and less accessible. Neuromorphic computing — building hardware that mimics the brain's sparse, event-driven computation — could close this gap by orders of magnitude.

The Robustness Gap

Current AI systems are remarkably brittle. A self-driving car might be confused by a stop sign partially obscured by a tree branch. An image classifier might be fooled by adding imperceptible noise to a photo. A language model might generate confident-sounding nonsense when asked about topics outside its training data.

Biological brains are far more robust. Humans can recognize a friend in a crowd despite bad lighting, unusual angles, and partial occlusion. We can navigate familiar environments even when they've been significantly rearranged. We can understand speech even with background noise, accents, and incomplete sentences.

This robustness comes from multiple sources: hierarchical processing, feedback loops, neuromodulation, and the ability to form internal models of the world. Current AI lacks all of these features.

The Adaptability Gap

Humans can learn new tasks with minimal instruction. Show a child a new game once, and they can often play it immediately. A chef can adapt a recipe based on what ingredients are available. A mechanic can troubleshoot a problem they've never seen before.

Current AI requires extensive training for each new task. A language model trained on English text can't suddenly understand Chinese without retraining. An image classifier trained on cats can't recognize dogs without seeing examples. This inflexibility makes current AI impractical for many real-world applications.

NeuroAI aims to close this adaptability gap by incorporating mechanisms like developmental learning, continual learning, and meta-learning — approaches that allow biological brains to learn new tasks quickly from few examples.

22. Why This Paper Matters

A Wake-Up Call for the AI Community

This paper is ultimately a wake-up call. The AI community has achieved remarkable things with deep learning, but those achievements have created a dangerous complacency. Because current AI works so well on many tasks, there's a temptation to believe that it's on the path to general intelligence — that we just need bigger models and more data.

The paper argues this is a mistake. Current AI is fundamentally limited in ways that can't be fixed by scaling. It lacks a body, a developmental process, feedback loops, neuromodulation, and the hierarchical organization that makes biological intelligence work. Adding more parameters won't create these features.

NeuroAI offers an alternative path — one that takes the brain's design principles seriously and builds AI from the ground up using those principles. This path is harder, slower, and less likely to produce quick commercial results. But it's the only path that can lead to AI that truly understands the world.

Why You Should Care

If you're not a scientist, you might wonder why this matters to you. The answer is simple: the future of AI will affect everyone. If current AI remains dominant, we'll get systems that are powerful but fragile, efficient at narrow tasks but unable to adapt, impressive in demos but unreliable in practice.

If NeuroAI succeeds, we'll get technology that is robust, adaptable, and efficient. We'll get robots that can safely work alongside humans, medical devices that restore lost abilities, and educational tools that genuinely help people learn. We'll get AI that enhances human cognition rather than replacing it.

The choice between these futures depends on the investments we make today. By understanding and supporting NeuroAI research, you're helping to shape the future of technology — and the future of humanity.

A Final Metaphor

Consider two kinds of airplanes. The first is a modern jet: fast, powerful, and impressive, but requiring enormous amounts of fuel, complex infrastructure, and expensive maintenance. The second is a bird: graceful, efficient, and adaptable, able to take off from almost anywhere and navigate complex environments with ease.

Current AI is like the jet — it achieves impressive feats through brute force. NeuroAI aims to build something more like the bird — technology that is not just powerful, but elegant, efficient, and truly intelligent.

The brain is nature's proof that such intelligence is possible. NeuroAI is the path to building technology that matches — and eventually surpasses — what evolution has achieved.

The Ultimate Question: Do we want technology that merely processes information, or technology that truly understands? NeuroAI offers a path to the latter. The journey will be long, but the destination is worth it.

23. Glossary of Key Terms

NeuroAI

A field that combines neuroscience and AI to build better artificial intelligence by learning from how the brain works.

Embodiment

The idea that intelligence requires a body that interacts with the physical world, not just software processing text and images.

Neuromorphic Engineering

Building computer hardware that mimics the structure and function of biological neural circuits, using artificial neurons and synapses.

LLM (Large Language Model)

AI systems like ChatGPT trained on massive text data to predict the next word in a sequence.

Connectome

A complete map of the neural connections in a brain - the wiring diagram of the nervous system.

Catastrophic Forgetting

The tendency of neural networks to forget previously learned information when trained on new data.

Neuromodulation

Chemical signals in the brain (like dopamine) that regulate which connections get strengthened or weakened during learning.

Praxicon

A hypothetical motor dictionary that stores movement patterns and their associated sensory inputs - the action equivalent of a word dictionary.

Spiking Neural Networks

Neural networks that communicate through discrete electrical spikes, like biological neurons, rather than continuous values.

Hierarchical Control

A system with multiple layers of control operating at different time scales - from fast reflexes to slow planning.

Von Neumann Architecture

The standard computer design where memory and processing are separate, requiring data to be shuttled back and forth.

AGI (Artificial General Intelligence)

Hypothetical AI with human-level intelligence across all cognitive tasks.

Neural Processing Unit (NPU)

Specialized AI chips built into modern smartphones for efficient on-device AI processing.

Flynn Effect

The observed rise in IQ scores over the 20th century, which has recently reversed (the anti-Flynn effect).

Swarm Intelligence

Collective behavior of decentralized, self-organized systems, like ant colonies or bird flocks.

Brain-Machine Interface (BMI)

A device that translates neural signals into commands for controlling external hardware like computers or robotic arms.

Deep Learning

A subset of machine learning using neural networks with many layers to learn hierarchical representations of data.

Reinforcement Learning

A type of machine learning where agents learn by receiving rewards or penalties for their actions.

Transfer Learning

Applying knowledge learned from one task to a different but related task.

Few-Shot Learning

Machine learning approaches that can learn from very few examples, similar to how humans learn quickly from limited data.

Event-Driven Computing

A computation paradigm where processing occurs only when new information arrives, rather than at fixed clock intervals.

Analog VLSI

Integrated circuit technology that processes continuous-valued signals, similar to how biological neurons process information.

3D Chip Integration

Stacking multiple layers of memory and processing on a single chip to reduce data movement and energy consumption.

Sparse Computing

Computation that only processes active elements, ignoring zeros and unused portions of data, similar to how the brain works.

Memory Consolidation

The process of stabilizing memories after initial acquisition, occurring primarily during sleep in biological systems.

About This Simplification

This document simplifies the paper "NeuroAI and Beyond" (arXiv:2601.19955) by Fellous et al., originally 53 pages of technical workshop proceedings. The goal was to make the ideas accessible to a general audience without losing the essential content.

Simplified by: @muhamedfazalps7
Original paper: arXiv:2601.19955