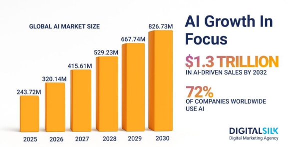
A Better Path Forward: Reconciling AI's Drive for Scale With the Reality of Its Energy Costs

Wiliam Fishell

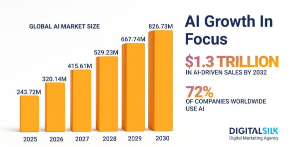
October 20th 2025

Overview

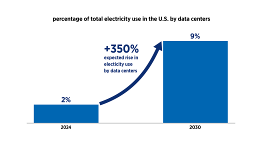


- The utilization of artificial intelligence products such as Large Language Models is **projected to grow 400% in the next 5 years**.

Overview

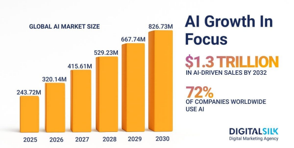


$\longrightarrow$

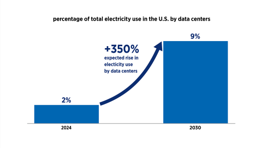


- The utilization of artificial intelligence products such as Large Language Models is **projected to grow 400% in the next 5 years**.
- In America alone this increased demand is expected to require almost **10% of all power by 2030: a 500% growth in AI's energy demand**.

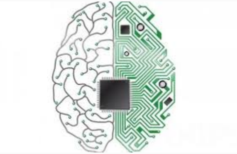
Overview



→

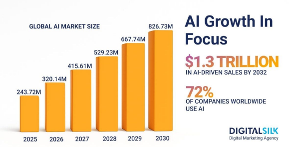


→

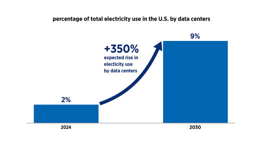


- The utilization of artificial intelligence products such as Large Language Models is **projected to grow 400% in the next 5 years**.
- In America alone this increased demand is expected to require almost **10% of all power by 2030: a 500% growth in AI's energy demand**.
- Looking to the **brain and biological principles offers a path forward** — enabling responsible scaling that is climate-conscious and equitable.

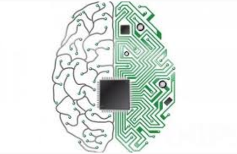
Overview



→



→



- The utilization of artificial intelligence products such as Large Language Models is **projected to grow 400% in the next 5 years**.
- In America alone this increased demand is expected to require almost **10% of all power by 2030: a 500% growth in AI's energy demand**.
- Looking to the **brain and biological principles offers a path forward** — enabling responsible scaling that is climate-conscious and equitable.

Columbia University in the City of New York

Harvard Medical School

Oak Ridge National Laboratory

Research conducted in collaboration with Columbia University, Harvard Medical School, and Oak Ridge National Laboratory.

Scaling Energy Infrastructure Today for AI Tomorrow

Meta's proposed $30B data center in Louisiana vs. the size of Manhattan

Microsoft secures nuclear power from the Three Mile Island reactor

$500B U.S. backed 'Stargate' data center initiative

- Demand from AI is **delaying modernization of the U.S. power grid**, forcing utilities to keep coal-fired plants online rather than retire them [1].

Scaling Energy Infrastructure Today for AI Tomorrow

Meta's proposed $30B data center in Louisiana vs. the size of Manhattan

Microsoft secures nuclear power from the Three Mile Island reactor

$500B U.S. backed 'Stargate' data center initiative

- Demand from AI is **delaying modernization of the U.S. power grid**, forcing utilities to keep coal-fired plants online rather than retire them [1].
- **Big Tech companies are delaying their net-zero carbon emission goals and are projected to miss future targets** [2].

Scaling Energy Infrastructure Today for AI Tomorrow

Meta's proposed $30B data center in Louisiana vs. the size of Manhattan

Microsoft secures nuclear power from the Three Mile Island reactor

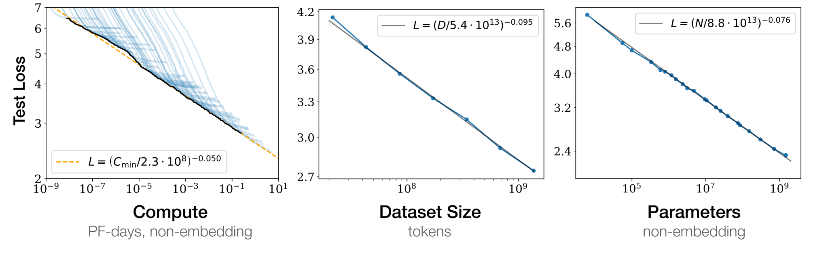
$500B U.S. backed 'Stargate' data center initiative

- Demand from AI is **delaying modernization of the U.S. power grid**, forcing utilities to keep coal-fired plants online rather than retire them [1].
- **Big Tech companies are delaying their net-zero carbon emission goals and are projected to miss future targets** [2].

Why is all of this power even necessary?

Bigger Models Get Better Results

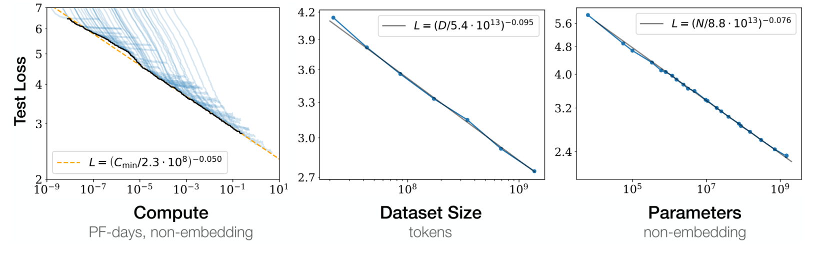
Models performance **scales logarithmically** with compute, dataset size and model-size



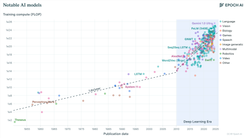
Model performance of neural networks w.r.t. **compute, dataset size, and model size** [3]

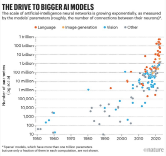
Bigger Models Get Better Results

Models performance **scales logarithmically** with compute, dataset size and model-size



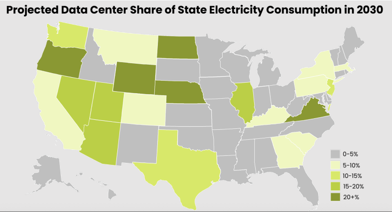
Model performance of neural networks w.r.t. **compute, dataset size, and model size** [3]



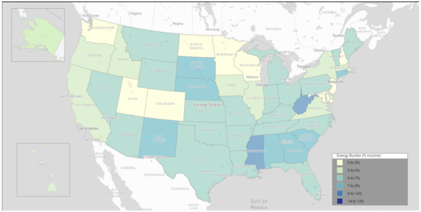


Companies **exponentially increasing the the size of these models and the compute used** to train them to keep up with scaling laws [4]

The burdens of AI Energy Demand Disproportionally Hurt Low Income Areas



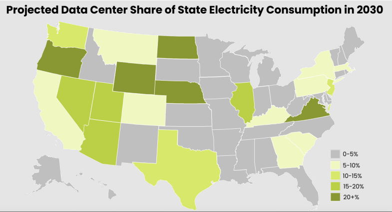
Share of Electricity Going to Data Centers by State



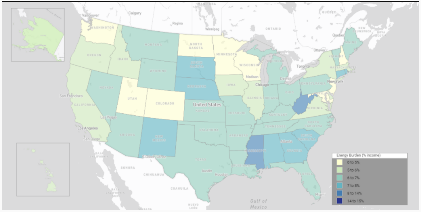
Share of Income per Household going to electricity

- Data center investment and energy consumption is **primarily concentrated in rural areas causing rising utilities prices**

The burdens of AI Energy Demand Disproportionally Hurt Low Income Areas



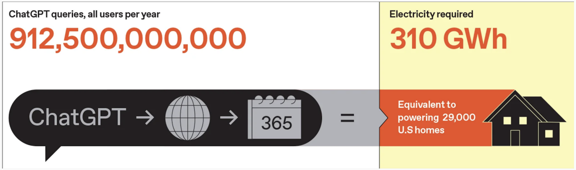
Share of Electricity Going to Data Centers by State



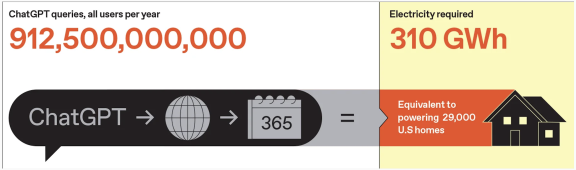
Share of Income per Household going to electricity

- Data center investment and energy consumption is **primarily concentrated in rural areas causing rising utilities prices**
- **Rising prices are not coupled with positive investment** for these local communities

What is Being Done To Address the Externalities of the AI Boom?

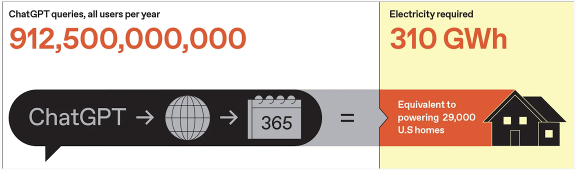


What is Being Done To Address the Externalities of the AI Boom?



Very little

What is Being Done To Address the Externalities of the AI Boom?

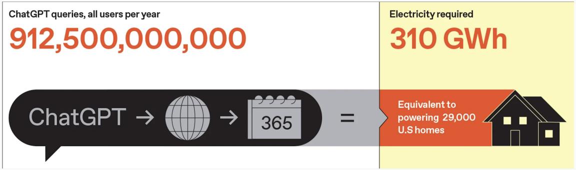


Very little

1. "Eventually, the cost of intelligence... will converge with the cost of energy." —Sam Altman

.

What is Being Done To Address the Externalities of the AI Boom?



Very little

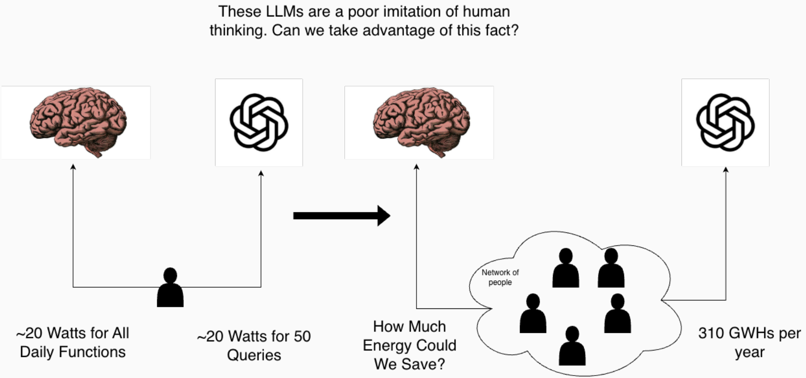
1. "Eventually, the cost of intelligence... will converge with the cost of energy." —Sam Altman
2. The Trump administration's EPA is actively rekindling U.S. fossil fuel production to fuel an "AI arms race with China" [5].

What is Being Done To Address the Externalities of the AI Boom?

"We are living proof that programs we want LLMs to do are possible" –Terry Sejnowski

What is Being Done To Address the Externalities of the AI Boom?

"We are living proof that programs we want LLMs to do are possible" –Terry Sejnowski



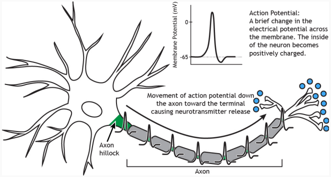
Can **biology** guide an answer to the side-effects of **the AI revolution?**

What Can Neuroscience Tell Us About Computation?

Neuromorphic Computing imitates 3 aspects of our brains **for more efficient computation**

What Can Neuroscience Tell Us About Computation?

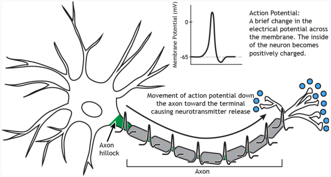
Neuromorphic Computing imitates 3 aspects of our brains **for more efficient computation**

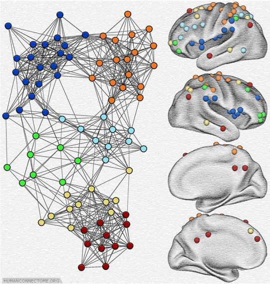


- **Action potential**: A way for encoding complex computation with a sparse set of signals

What Can Neuroscience Tell Us About Computation?

Neuromorphic Computing imitates 3 aspects of our brains **for more efficient computation**

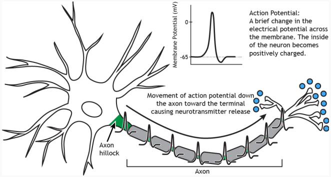


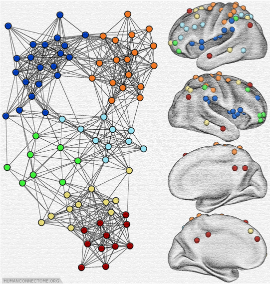


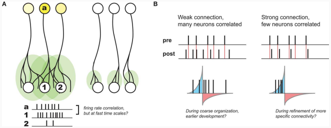
- **Action potential**: A way for encoding complex computation with a sparse set of signals
- **Massive Parallelization**: The connection of multiple neural pathways enabling for a large amount of multi tasking

What Can Neuroscience Tell Us About Computation?

Neuromorphic Computing imitates 3 aspects of our brains **for more efficient computation**



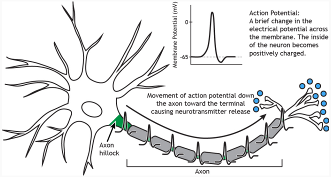


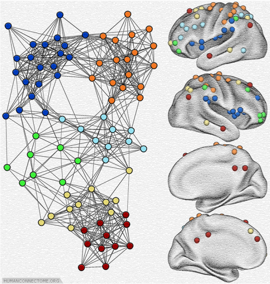


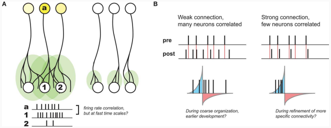
- **Action potential**: A way for encoding complex computation with a sparse set of signals
- **Massive Parallelization**: The connection of multiple neural pathways enabling for a large amount of multi tasking
- **Spike Time Dependent Plasticity**: "Neurons that fire together wire together"; A biologically plausible way of learning

What Can Neuroscience Tell Us About Computation?

Neuromorphic Computing imitates 3 aspects of our brains **for more efficient computation**



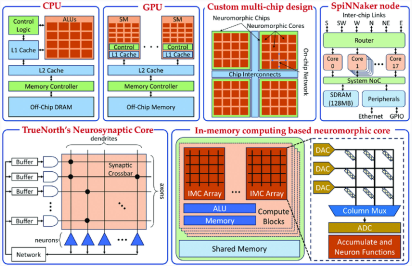




- **Action potential**: A way for encoding complex computation with a sparse set of signals
- **Massive Parallelization**: The connection of multiple neural pathways enabling for a large amount of multi tasking
- **Spike Time Dependent Plasticity**: "Neurons that fire together wire together"; A biologically plausible way of learning

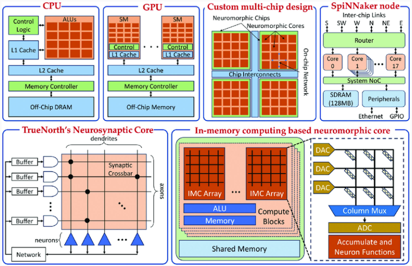
What happens when we build software and Hardware Based on these Processes?

Neuromorphic Computing is an Ultra-Efficient, Brain-Inspired Approach for Building Hardware and Software

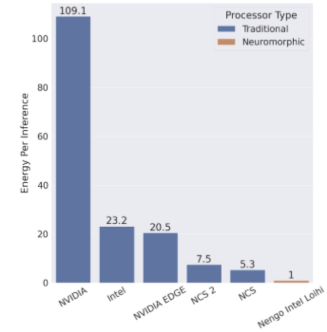


- **CPUs are not used for machine learning** because they do things in series, making learning very slow
- **GPUs are the backbone of machine learning because of their parallelized architecture**; the dominance of GPUs is why NVIDIA is worth so much money
- **Neuromorphic Chips are customized chips that simulate key neuroscience principles** and enable massive parallelization

Neuromorphic Computing is an Ultra-Efficient, Brain-Inspired Approach for Building Hardware and Software

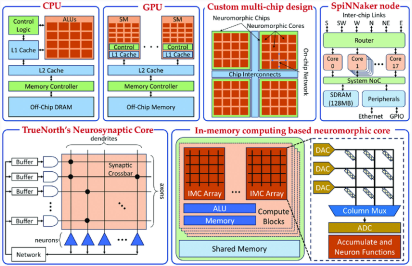


GPU vs. Neuromorphic

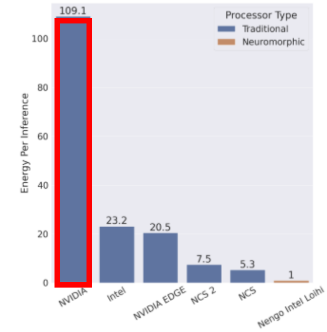


- **CPUs are not used for machine learning** because they do things in series, making learning very slow
- **GPUs are the backbone of machine learning because of their parallelized architecture**; the dominance of GPUs is why NVIDIA is worth so much money
- **Neuromorphic Chips are customized chips that simulate key neuroscience principles** and enable massive parallelization

Neuromorphic Computing is an Ultra-Efficient, Brain-Inspired Approach for Building Hardware and Software

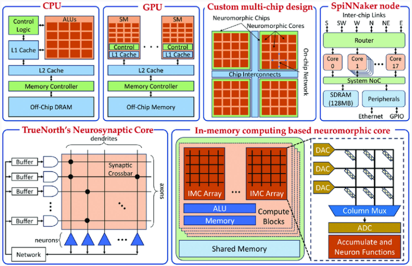


GPU vs. Neuromorphic

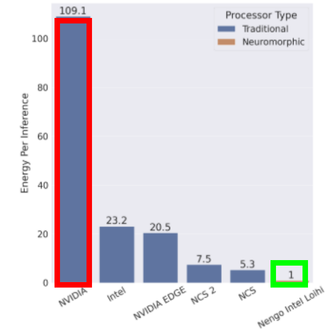


- **CPUs are not used for machine learning** because they do things in series, making learning very slow
- **GPUs are the backbone of machine learning because of their parallelized architecture**; the dominance of GPUs is why NVIDIA is worth so much money
- **Neuromorphic Chips are customized chips that simulate key neuroscience principles** and enable massive parallelization

Neuromorphic Computing is an Ultra-Efficient, Brain-Inspired Approach for Building Hardware and Software

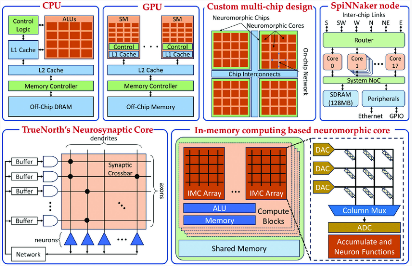


GPU vs. Neuromorphic

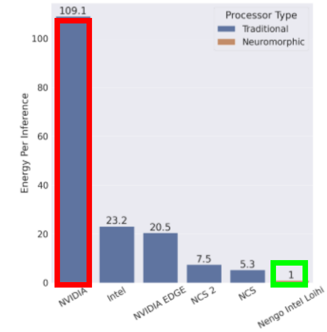


- **CPUs are not used for machine learning** because they do things in series, making learning very slow
- **GPUs are the backbone of machine learning because of their parallelized architecture**; the dominance of GPUs is why NVIDIA is worth so much money
- **Neuromorphic Chips are customized chips that simulate key neuroscience principles** and enable massive parallelization

Neuromorphic Computing is an Ultra-Efficient, Brain-Inspired Approach for Building Hardware and Software



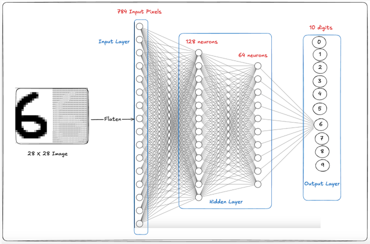
GPU vs. Neuromorphic



- **CPUs are not used for machine learning** because they do things in series, making learning very slow
- **GPUs are the backbone of machine learning because of their parallelized architecture**; the dominance of GPUs is why NVIDIA is worth so much money
- **Neuromorphic Chips are customized chips that simulate key neuroscience principles** and enable massive parallelization

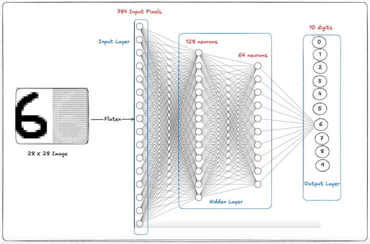
If neuromorphic hardware is so **much more power efficient while maintaining performance**, what obstacles are preventing widespread adoption of this technology?

Neuromorphic Hardware Delivers its Energy Efficiency Only When Paired With Biologically Inspired Algorithms

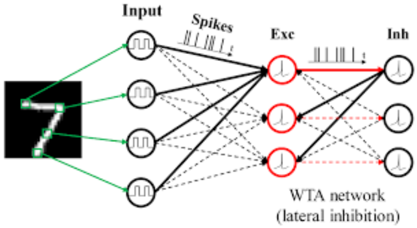


Classic Neural Network Architecture For Classifying Handwritten Digits

Neuromorphic Hardware Delivers its Energy Efficiency Only When Paired With Biologically Inspired Algorithms

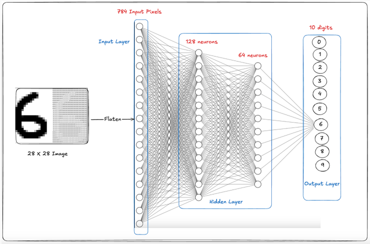


Classic Neural Network Architecture For Classifying Handwritten Digits

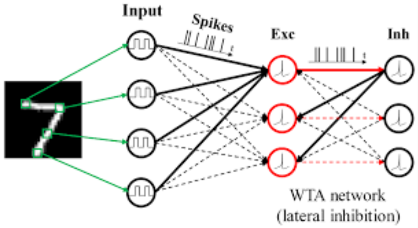


Spiking Neural Network Architecture For Classifying Handwritten Digits

Neuromorphic Hardware Delivers its Energy Efficiency Only When Paired With Biologically Inspired Algorithms



Classic Neural Network Architecture For Classifying Handwritten Digits

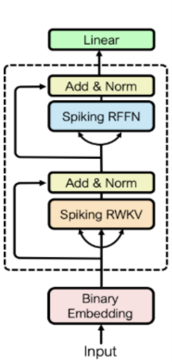


Spiking Neural Network Architecture For Classifying Handwritten Digits

- A normal neural network treats **each pixel as an input neuron**
- Spiking neural networks, inspired by biology, **use rate encoding schemes to represent the same information**

Despite Limited Use, Spiking Neural Networks Have Demonstrated Promising Results in LLMs

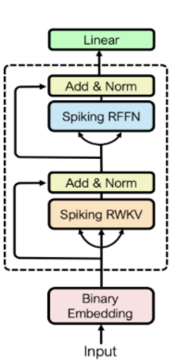
SpikeGPT



- SpikeGPT matched GPT-3's performance with **10× fewer parameters and 33× less power during inference** [6]

Despite Limited Use, Spiking Neural Networks Have Demonstrated Promising Results in LLMs

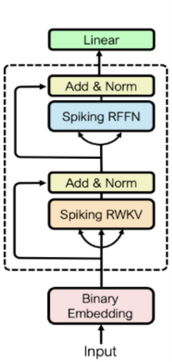
SpikeGPT



- SpikeGPT matched GPT-3's performance with **10× fewer parameters and 33× less power during inference** [6]
- SNNs have an **inductive bias towards low-frequency functions**, generally making them **not optimal for modeling linguistic relationships** [7]

Despite Limited Use, Spiking Neural Networks Have Demonstrated Promising Results in LLMs

SpikeGPT

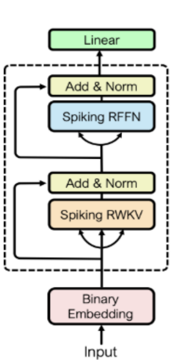


- SpikeGPT matched GPT-3's performance with **10× fewer parameters and 33× less power during inference** [6]
- SNNs have an **inductive bias towards low-frequency functions**, generally making them **not optimal for modeling linguistic relationships** [7]

If this works so well, where is the Neuromorphic revolution?

Despite Limited Use, Spiking Neural Networks Have Demonstrated Promising Results in LLMs

SpikeGPT



- SpikeGPT matched GPT-3's performance with **10× fewer parameters and 33× less power during inference** [6]
- SNNs have an **inductive bias towards low-frequency functions**, generally making them **not optimal for modeling linguistic relationships** [7]

If this works so well, where is the Neuromorphic revolution?

These neuromorphic language models struggle at modeling long-range dependencies **without a lot** of training data

Traditional Language Models (Transformers) Are Able to Model Long-Range Dependencies With Few Examples

Long-Range Dependencies are relationships in sequential data where predictions are **dependent** on earlier points

Traditional Language Models (Transformers) Are Able to Model Long-Range Dependencies With Few Examples

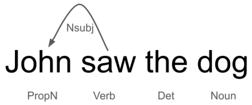
Long-Range Dependencies are relationships in sequential data where predictions are **dependent** on earlier points

Dependency Parsing in Language Modeling

Traditional Language Models (Transformers) Are Able to Model Long-Range Dependencies With Few Examples

Long-Range Dependencies are relationships in sequential data where predictions are **dependent** on earlier points

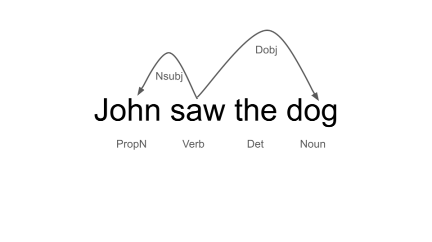
Dependency Parsing in Language Modeling



Traditional Language Models (Transformers) Are Able to Model Long-Range Dependencies With Few Examples

Long-Range Dependencies are relationships in sequential data where predictions are **dependent** on earlier points

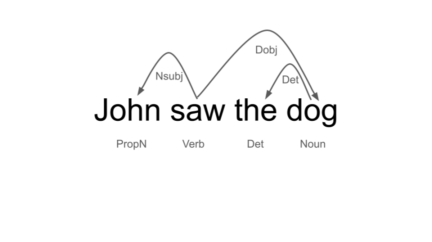
Dependency Parsing in Language Modeling



Traditional Language Models (Transformers) Are Able to Model Long-Range Dependencies With Few Examples

Long-Range Dependencies are relationships in sequential data where predictions are **dependent** on earlier points

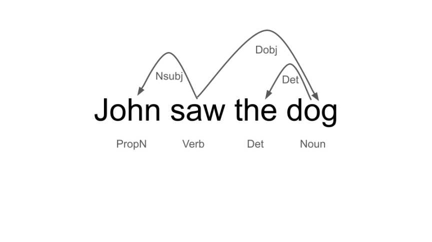
Dependency Parsing in Language Modeling



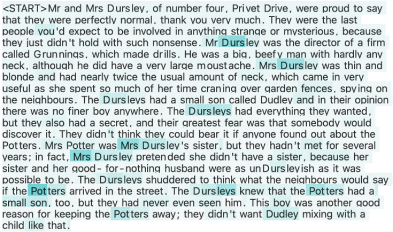
Traditional Language Models (Transformers) Are Able to Model Long-Range Dependencies With Few Examples

Long-Range Dependencies are relationships in sequential data where predictions are **dependent** on earlier points

Dependency Parsing in Language Modeling

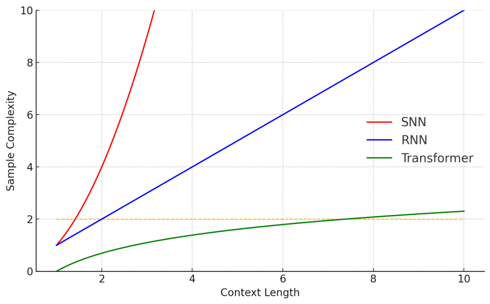


Word Associations in Harry Potter Learned by LLM



Unbounded Excitation in Spiking Neurons Causes Poor Learning of Long Contexts in Language Modeling Tasks

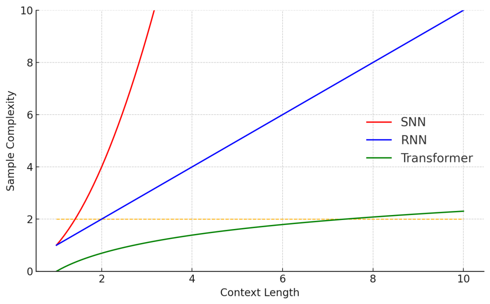
Samples Needed For PAC Learnability For Popular Language Modeling Architectures



This work was recently submitted to BioRxiv and presented at the SHDA Workshop at the International Conference on Supercomputing (ICS 2025).

Unbounded Excitation in Spiking Neurons Causes Poor Learning of Long Contexts in Language Modeling Tasks

Samples Needed For PAC Learnability For Popular Language Modeling Architectures

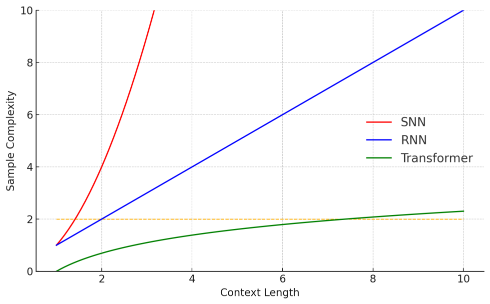


- As context length grows longer, **neuronal activity increases causing nosier and nosier networks**

This work was recently submitted to BioRxiv and presented at the SHDA Workshop at the International Conference on Supercomputing (ICS 2025).

Unbounded Excitation in Spiking Neurons Causes Poor Learning of Long Contexts in Language Modeling Tasks

Samples Needed For PAC Learnability For Popular Language Modeling Architectures

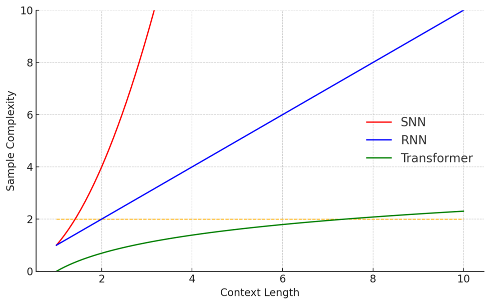


- As context length grows longer, **neuronal activity increases causing nosier and nosier networks**
- **SNNs don't have normilzation methods like self-attention**

This work was recently submitted to BioRxiv and presented at the SHDA Workshop at the International Conference on Supercomputing (ICS 2025).

Unbounded Excitation in Spiking Neurons Causes Poor Learning of Long Contexts in Language Modeling Tasks

Samples Needed For PAC Learnability For Popular Language Modeling Architectures



- As context length grows longer, **neuronal activity increases causing nosier and nosier networks**
- **SNNs don't have normilzation methods like self-attention**

Biology can save the day once again!

This work was recently submitted to BioRxiv and presented at the SHDA Workshop at the International Conference on Supercomputing (ICS 2025).

We Are Currently Investigating How Mimicking Inhibitory Neurons Can Help SNNs Learn Long-Range Dependencies

Lateral inhibitory processes in real neurons are used to **bound endless excitation**

We Are Currently Investigating How Mimicking Inhibitory Neurons Can Help SNNs Learn Long-Range Dependencies

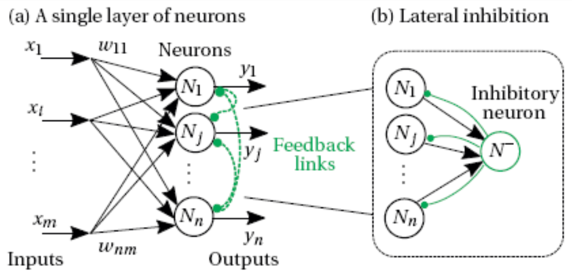
Lateral inhibitory processes in real neurons are used to **bound endless excitation**

Due to its inducing of sparse networks, **we believe inhibition is crucial to how biological systems model long-range dependencies**

We Are Currently Investigating How Mimicking Inhibitory Neurons Can Help SNNs Learn Long-Range Dependencies

Lateral inhibitory processes in real neurons are used to **bound endless excitation**

Due to its inducing of sparse networks, **we believe inhibition is crucial to how biological systems model long-range dependencies**

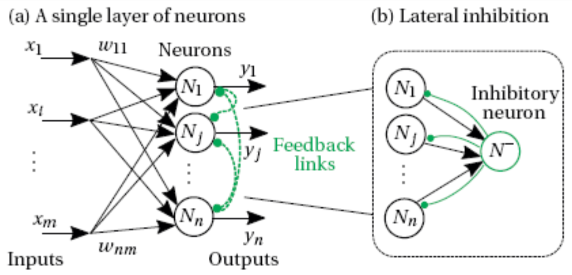


1. When a neuron spikes it sends inhibitory signals to other neurons in the same layer
2. **Inhibitory layers** can be used to **reduce neuronal activity** helping denoise networks

We Are Currently Investigating How Mimicking Inhibitory Neurons Can Help SNNs Learn Long-Range Dependencies

Lateral inhibitory processes in real neurons are used to **bound endless excitation**

Due to its inducing of sparse networks, **we believe inhibition is crucial to how biological systems model long-range dependencies**

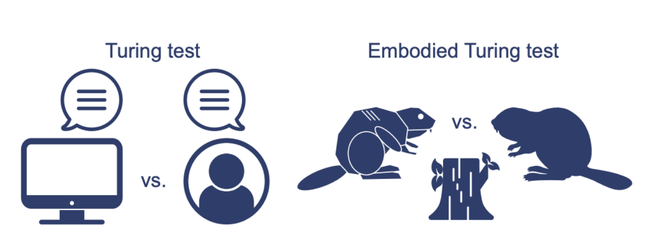


1. When a neuron spikes it sends inhibitory signals to other neurons in the same layer
2. **Inhibitory layers** can be used to **reduce neuronal activity** helping denoise networks

Where else does Neuromorphic computing fit into the AI revolution?

Artificial Intelligence Exists Beyond Just Language Models

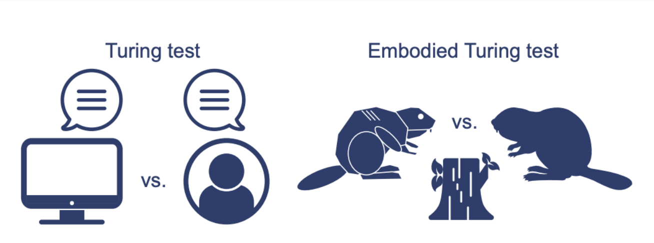
Embodied Intelligence is the next frontier for Machine Learning



- **The Turing test says a system is intelligent if a human can't tell whether it's a machine or a person**
- Generally it is understood that LLMs have solved this problem

Artificial Intelligence Exists Beyond Just Language Models

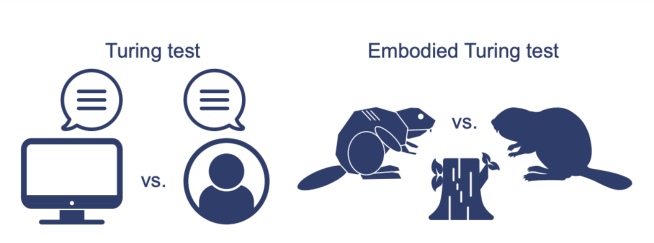
Embodied Intelligence is the next frontier for Machine Learning



- **The Turing test says a system is intelligent if a human can't tell whether it's a machine or a person**
- Generally it is understood that LLMs have solved this problem
- **The Embodied Turing test states that An AI animal model—whether robotic or in simulation—passes the test if its behavior is indistinguishable from that of its living counterpart** [8]
- **This problem remains unsolved**

Artificial Intelligence Exists Beyond Just Language Models

Embodied Intelligence is the next frontier for Machine Learning

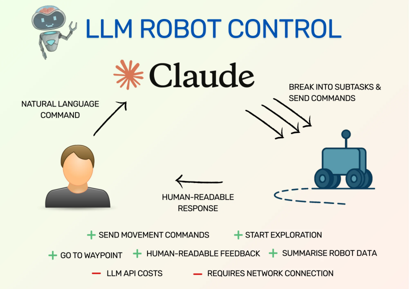


- **The Turing test says a system is intelligent if a human can't tell whether it's a machine or a person**
- Generally it is understood that LLMs have solved this problem
- **The Embodied Turing test states that An AI animal model—whether robotic or in simulation—passes the test if its behavior is indistinguishable from that of its living counterpart** [8]
- **This problem remains unsolved**

For our purposes we will focus on **autonomous driving for our Embodied Turing test**

Current Machine Learning Methods are Either Unsuitable for Robotics or not Scalable due to their Energy Demands

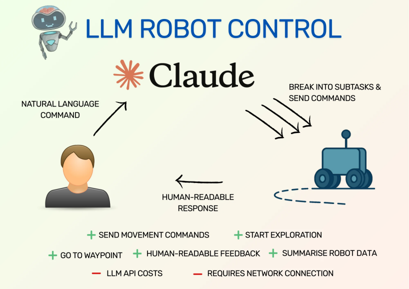
Robot Controlled via LLM



- Controlling a robot through an LLM is a **terrible idea**

Current Machine Learning Methods are Either Unsuitable for Robotics or not Scalable due to their Energy Demands

Robot Controlled via LLM

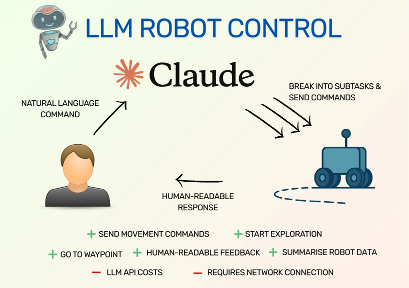


Vision Model Controlling Waymo

- Controlling a robot through an LLM is a **terrible idea**
- Embodied intelligence demands **on-chip computation and models that are fast, safe, and efficient**
- While systems like Waymo can operate effectively on today's chips, the **energy efficiency of neuromorphic hardware enables intelligence to scale far more**

Current Machine Learning Methods are Either Unsuitable for Robotics or not Scalable due to their Energy Demands

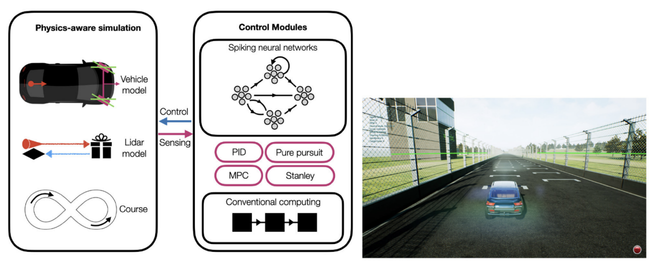
Robot Controlled via LLM



Vision Model Controlling Waymo

- Controlling a robot through an LLM is a **terrible idea**
- Embodied intelligence demands **on-chip computation and models that are fast, safe, and efficient**
- While systems like Waymo can operate effectively on today's chips, the **energy efficiency of neuromorphic hardware enables intelligence to scale far more**
- **Neuromorphic computing is a perfect substitute** for these types of embodied intelligence

Autonomous Driving is Already Turning to Neuromorphic Computing For Better Performance on Power Constrained Systems

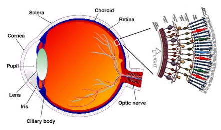


SNN architecture for autonomous driving

- Large automotive companies such as **Mercedes are actively researching Neuromorphic setups for the future of self-driving** [9]
- **Neuromorphic computing scales exceptionally well for vision tasks** — delivering performance comparable to classical vision models, but at only a fraction of the energy cost
- Autonomous driving is a leading example of how **neuromorphic systems are beginning to rival traditional architectures in embodied intelligence** [10]

Emulating Convolution Still Demands Many Synapses, Making Neuromorphic Models Computationally Expensive at Scale

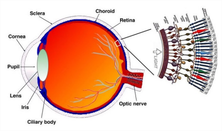
Neuronal Model of Eye



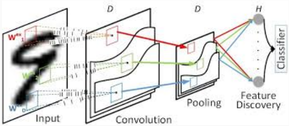
- Our **eyes convolve specific localities** using neurons that are hardwired to that location

Emulating Convolution Still Demands Many Synapses, Making Neuromorphic Models Computationally Expensive at Scale

Neuronal Model of Eye



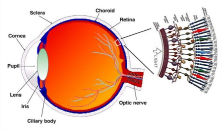
Spiking Convolution Model



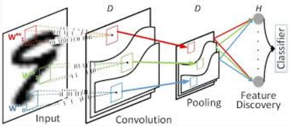
- Our **eyes convolve specific localities** using neurons that are hardwired to that location
- **SNNs mimic this hard wiring process** — but to detect a specific object, the same feature must be instantiated across many neurons, repeatedly and redundantly

Emulating Convolution Still Demands Many Synapses, Making Neuromorphic Models Computationally Expensive at Scale

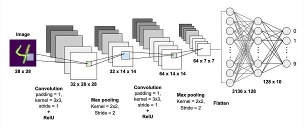
Neuronal Model of Eye



Spiking Convolution Model



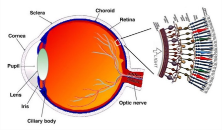
Traditional Convolutional Network



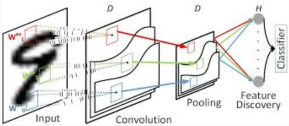
- Our **eyes convolve specific localities** using neurons that are hardwired to that location
- **SNNs mimic this hard wiring process** — but to detect a specific object, the same feature must be instantiated across many neurons, repeatedly and redundantly
- **Traditional convolutional networks do not mimic biology** so neurons are not hardwired in the same way avoiding redundancy

Emulating Convolution Still Demands Many Synapses, Making Neuromorphic Models Computationally Expensive at Scale

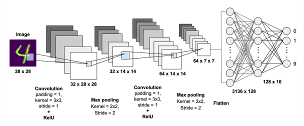
Neuronal Model of Eye



Spiking Convolution Model



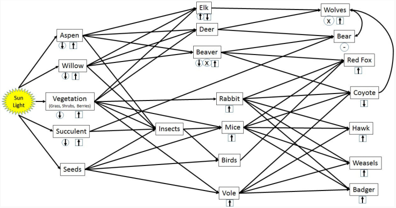
Traditional Convolutional Network



- Our **eyes convolve specific localities** using neurons that are hardwired to that location
- **SNNs mimic this hard wiring process** — but to detect a specific object, the same feature must be instantiated across many neurons, repeatedly and redundantly
- **Traditional convolutional networks do not mimic biology** so neurons are not hardwired in the same way avoiding redundancy

Biological plausibility offers powerful tools for designing more efficient machine learning architectures, but it **shouldn't be followed blindly**

Inspired by complex adaptive systems, we simulate neural circuits while reproducing convolution-like computation



Wolf Introduction To Yellow Stone National Park

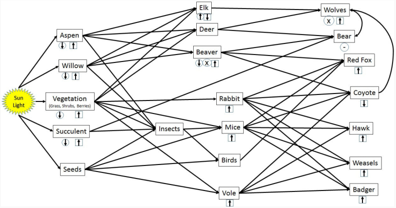
SAGE Sim
SCALABLE AGENT-BASED GPU ENABLED SIMULATOR

Agent based simulation software for modeling complex adaptive systems [11]

- SageSim enables **modeling of complex adaptive systems** through simulating individual agents in the system

This work is currently being completed by the Learning Systems Group at Oak Ridge National Laboratory.

Inspired by complex adaptive systems, we simulate neural circuits while reproducing convolution-like computation



Wolf Introduction To Yellow Stone National Park

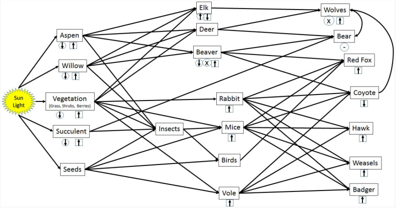
SAGE Sim
SCALABLE AGENT-BASED GPU ENABLED SIMULATOR

Agent based simulation software for modeling complex adaptive systems [11]

- SageSim enables **modeling of complex adaptive systems** through simulating individual agents in the system
- **Neuromorphic systems are simply complex adaptive systems**

This work is currently being completed by the Learning Systems Group at Oak Ridge National Laboratory.

Inspired by complex adaptive systems, we simulate neural circuits while reproducing convolution-like computation



Wolf Introduction To Yellow Stone National Park

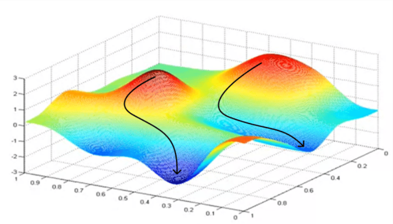
SAGESim
SCALABLE AGENT-BASED GPU ENABLED SIMULATOR

Agent based simulation software for modeling complex adaptive systems [11]

- SageSim enables **modeling of complex adaptive systems** through simulating individual agents in the system
- **Neuromorphic systems are simply complex adaptive systems**
- Convolutional networks can achieve neuromorphic-level energy efficiency—**without** the computational burden of **hard-wiring synapses—by modeling neurons and synapses through an agent-based architecture**

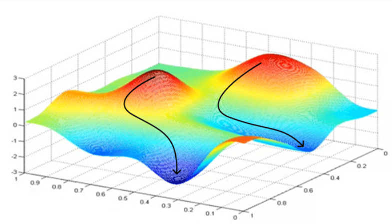
This work is currently being completed by the Learning Systems Group at Oak Ridge National Laboratory.

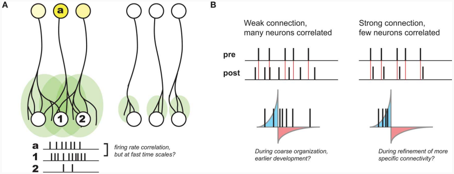
This Neuromorphic Convoltuion Architecture Enables Biologically Realistic Learning



- **Gradient descent** is the dominant mode of learning deep learning neural networks in traditional and Neuromorphic settings

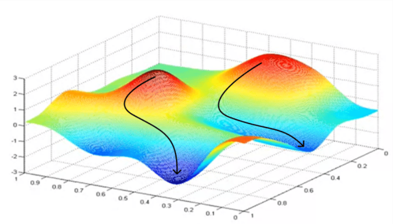
This Neuromorphic Convoltuion Architecture Enables Biologically Realistic Learning

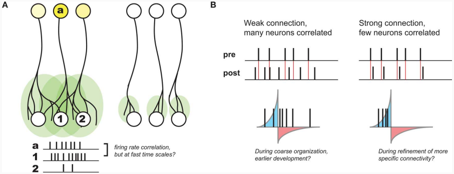




- **Gradient descent** is the dominant mode of learning deep learning neural networks in traditional and Neuromorphic settings
- By treating the **spiking convolution network as a complex adaptive system** we can use **spike-time dependent plasticity** to learn the network

This Neuromorphic Convoltuion Architecture Enables Biologically Realistic Learning

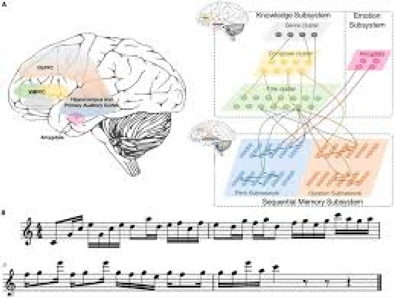




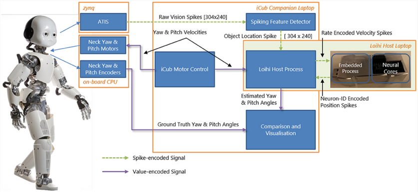
- **Gradient descent** is the dominant mode of learning deep learning neural networks in traditional and Neuromorphic settings
- By treating the **spiking convolution network as a complex adaptive system** we can use **spike-time dependent plasticity** to learn the network
- This is **more efficient** then traditional deep learning and enables **continuous learning** once the model is deployed

Neuromorphic Computing: A Promising but Not Yet Scalable Solution to AI's Energy Crisis

Neuromorphic computing presents a **better path** forward through biology as **current LLMs and robots power demands are unsustainable**



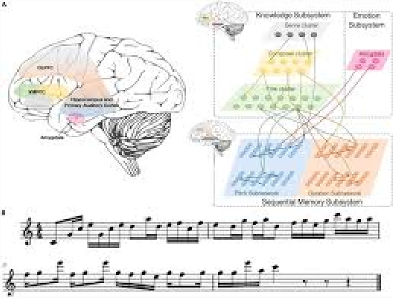
Biological representation of an SNN learning compositional language



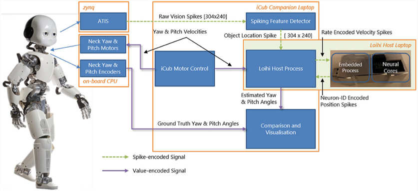
Humaniod Robot Utilizing Spiking and Neuromorphic Hardware

Neuromorphic Computing: A Promising but Not Yet Scalable Solution to AI's Energy Crisis

Neuromorphic computing presents a **better path** forward through biology as **current LLMs and robots power demands are unsustainable**



Biological representation of an SNN learning compositional language

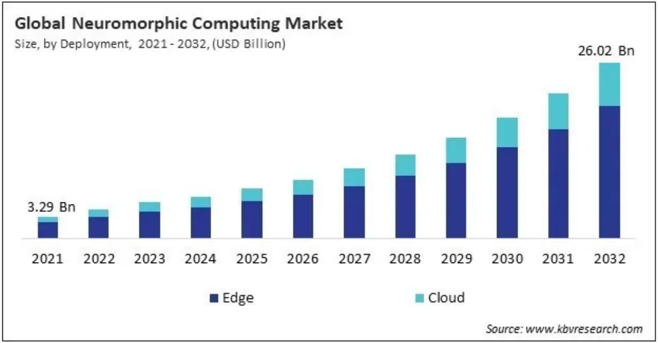


Humaniod Robot Utilizing Spiking and Neuromorphic Hardware

Despite all this hype, **Neuromorphic computing** is considered the nuclear fusion of computer science: it **over promises and under delivers**

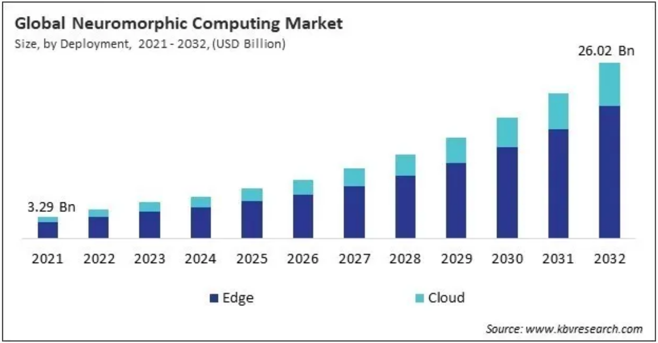
Neuromorphic Computing: A Promising but Not Yet Scalable Solution to AI's Energy Crisis

By **2030** global Neuromorphic hardware sales will represent **less then 7%** of global GPU sales



Neuromorphic Computing: A Promising but Not Yet Scalable Solution to AI's Energy Crisis

By **2030** global Neuromorphic hardware sales will represent **less then 7%** of global GPU sales



Neuromorphic computing still has **significant ground to cover** before it can become a viable competitor to **today's SOTA AI models and infrastructure**

What Can We Do To Bring to Effectuate This Change?

1. **Carbon taxes** that incentivize big tech companies to search for alternatives

What Can We Do To Bring to Effectuate This Change?

1. **Carbon taxes** that incentivize big tech companies to search for alternatives
2. **Push for better funding for this basic research**

What Can We Do To Bring to Effectuate This Change?

1. **Carbon taxes** that incentivize big tech companies to search for alternatives
2. **Push for better funding for this basic research**
3. Ultimately, we need to **convince an administration hostile to climate change** that it is in their best interest to **embrace this technology**

What Can We Do To Bring to Effectuate This Change?

1. **Carbon taxes** that incentivize big tech companies to search for alternatives
2. **Push for better funding for this basic research**
3. Ultimately, we need to **convince an administration hostile to climate change** that it is in their best interest to **embrace this technology**

Let's open this up to a **broader discussion** — not just questions **about the work itself, but also your thoughts on the societal, technical, and policy implications of neuromorphic computing**, and whether this is truly a direction we want to pursue and accelerate

Thank You For Listening!

L. Kearney and T. Gardner, "Us urges utilities to keep coal-fired plants running as ai demand booms, energy secretary says," *Reuters*, Sep. 25 2025, accessed: 2025-10-18. [Online]. Available: https://www.reuters.com/sustainability/climate-energy/us-urges-utilities-keep-coal-fired-plants-running-ai-demand-booms-en

C. Brown, "Tech companies' climate goals face growing pressure," *The New York Times*, Aug. 05 2025, accessed: 2025-10-18. [Online]. Available: https://www.nytimes.com/2025/08/05/climate/tech-companies-climate-goals.html

J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020.

J. Sevilla, L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn, and P. Villalobos, "Compute trends across three eras of machine learning," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–8.

J. Bolster. (2025, Sep.) Citing 'ai arms race,' trump administration announces efforts to rekindle us coal industry. Inside Climate News. Accessed: 2025-10-18. [Online]. Available: https://insideclimatenews.org/news/29092025/trump-administration-announces-coal-industry-efforts/

R.-J. Zhu, Q. Zhao, G. Li, and J. K. Eshraghian, "Spikegpt: Generative pre-trained language model with spiking neural networks," *arXiv preprint arXiv:2302.13939*, 2023.

W. Dorrell, M. Yuffa, and P. E. Latham, "Meta-learning the inductive bias of simple neural circuits," in *International*

Conference on Machine Learning. PMLR, 2023, pp. 8389–8402.

A. Zador, S. Escola, B. Richards, B. Ölveczky, Y. Bengio, K. Boahen, M. Botvinick, D. Chklovskii, A. Churchland, C. Clopath *et al.*, "Catalyzing next-generation artificial intelligence through neuroai," *Nature communications*, vol. 14, no. 1, p. 1597, 2023.

Mercedes-Benz Group AG, "Neuromorphic computing: More energy efficiency in autonomous driving of the future," https://group.mercedes-benz.com/innovations/product-innovation/autonomous-driving/neuromorphic-computing.html, Dec. 2024, accessed: YYYY-MM-DD.

R. Halaly and E. Ezra Tsur, "Autonomous driving controllers with neuromorphic spiking neural networks," *Frontiers in Neurorobotics*, vol. 17, p. 1234962, 2023.

C. Gunarathne, X. Zhang, R. Patton, and M. Coletti, "Sagesim: Scalable agent-based gpu-enabled simulator," U.S. Department of Energy (DOE) Software, 2024, distributed GPU agent-based modeling framework in Python. [Online]. Available: https://doi.org/10.11578/DC.20240327.1