

Contributed Article for: AiThORITY

Brent Oster: Founder and CEO of [ORBAI](#)

Practical Tips for Developing an Artificial General Intelligence

What we usually think of as Artificial Intelligence (AI) today, when we see human-like robots and holograms in our fiction, talking and acting like real people and having human-level or even superhuman intelligence and capabilities, is actually called Artificial General Intelligence (AGI), and it does NOT exist anywhere on earth yet. What actually have for AI today is much simpler and much more narrow Deep Learning (DL) that can only do some very specific tasks better than people. It has fundamental limitations that will not allow it to become AGI, so if that is our goal, we need to innovate and come up with better networks and better methods for shaping them into an artificial brain.

Classically human tasks such as writing, speaking, making sense of new ideas and situations remain far outside the realm of contemporary AI. For example, DL uses deep 'neural' networks (DNNs) that really have very little in common with biological neurons. They are just summation units with an activation function feeding a static number to connections that instantly 'communicate' that number to all the 'neurons' in the next layer, each modulated by the weight of that connection. The high-level architecture defined by types and numbers of neurons is defined by a human, often through a laborious process of trial and error without a comprehensive understanding of why certain layers and structures are effective. Only the connection weights can be modified, and for training a DNN we use a technique called backpropagation to iteratively adjust the weights such that the output matches what we expect (given the input), meaning we have to label our training data with the expected outputs ahead of time, do static training with it, and DNNs cannot generally learn dynamically in deployment. These techniques are used in convolutional neural networks (CNNs) for image data and recurrent neural networks (RNNs) for sequential data such as language. Both CNNs and RNNs are combined in a composite network to process video and speech.

The main shortcomings to Deep Neural Network based AI today:

- DNNs do static evaluation of an input to an output
- Do not hold state, do processing or operations beyond summation at the layers
- They require a large number of layers to do a useful task
- They require large amounts of labelled training data
- They generalize poorly to similar problems outside the scope of the training data
- They use enormous computational resources and time in training
- They cannot handle temporal-spatial operations (beyond a very simple level with RNNs)
- The speech interfaces today using DL are far from being conversational, let alone at a human level.
- DNNs, CNNs, RNNs, ... are all hacks, simplified shadows of biological neural functioning,
- No matter how large they scale, they will never result in an AGI

What we need on the road to AGI are better neural networks to start with. The human brain is a very sophisticated bio-chemical-electrical computer with around 100 billion neurons and 100 trillion connections (and synapses) between them. I will describe two decades of neuroscience in the next two paragraphs, so if you trust me so far, feel free to skip. Here are two good videos about the biological [Neuron](#) and [Synapse](#) from '2-Minute Neuroscience' on YouTube that will also help.

Each neuron takes in spikes of electrical charge from its dendrites and performs a very complicated integration in time and space, resulting in the charge accumulating in the neuron and (once exceeding an action potential) causing the neuron to fire spikes of electricity out along its axon, moving in time and space as that axon branches and re-amplifies the signal, carrying it to thousands of synapses, where it is absorbed by each synapse. This process causes neurotransmitters to be emitted into the synaptic cleft, where they are chemically integrated (with ambient neurochemistry contributing). These neurotransmitters migrate across the cleft to the post-synaptic side, where their accumulation in various receptors eventually cause the post-synaptic side to fire a spike down along the dendrite to the next neuron. When a spike enters a synapse and is emitted out the other side within a certain time, that synapse becomes more sensitive or potentiated, and fires more easily. We call this Hebbian learning, which is constantly occurring as we move around and interact with our environment.

The brain is organized into cortices for processing sensory inputs, motor control, language understanding, language creation, speaking and cognition and planning, and logic. Each of these cortices has very sophisticated space and time signal processing, including feedback loops and bidirectional networks, so visual input is processed into abstractions or 'thoughts' by one directional network, and then those thoughts are processed back out to a recreation of the expected visual representation by another, complementary network in the opposite direction, and they are fed back into each other throughout. Picture a 'fire truck' with your eyes closed and you will see the feedback network of your visual cortex at work, allowing you to visualize the 'thought' of a fire truck into an image of one. You could probably even draw it if you wanted.

[Predictive Coding in the Visual Cortex – Rao, Ballard](#)

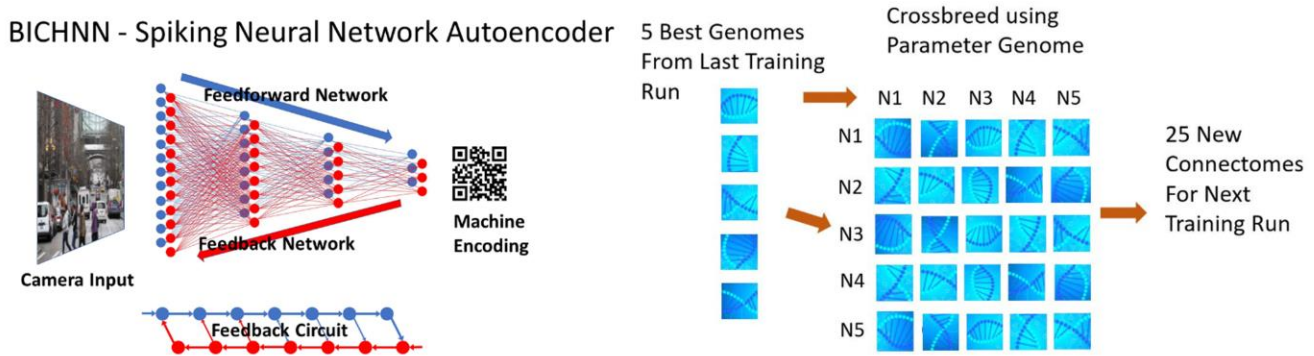
These feedback loops train our sensory cortices to encode the information from our senses into compact 'thoughts' the rest of the brain can use, as well as providing a perceptual filter by comparing what we are seeing to what we expect to see, so our visual cortex can focus on what we are looking for and screen the rest out. The frontal and pre-frontal cortex have tighter, more specialized feedback loops that can store state (short-term memory), operate on it, and perform logic and planning at the macroscale. All our cortices work together and can learn associatively and store long-term memories by [Hebbian learning](#).

Neuromorphic computing uses spiking neural networks (SNNs), which model neurons as discrete computational units that work much more like biological neurons, fundamentally computing in the time domain, approximating them with simple models like [Izhikevich](#) or more complex ones like [Hodgkin-Huxley](#) (Nobel Prize 1953). However, to date, application of spiking neural networks has remained difficult, as finding a way to train them to do specific tasks has remained elusive. Although Hebbian learning functions in these networks, there has not been a way to shape them so we can train them to learn to do specific tasks. Backpropagation (used in DNNs) does not work because all these spiking signals are one-way in time and are emitted, absorbed and integrated in operations that are non-reversible.

Article: [Spiking Neural Networks - the Next Generation of Machine Learning](#)

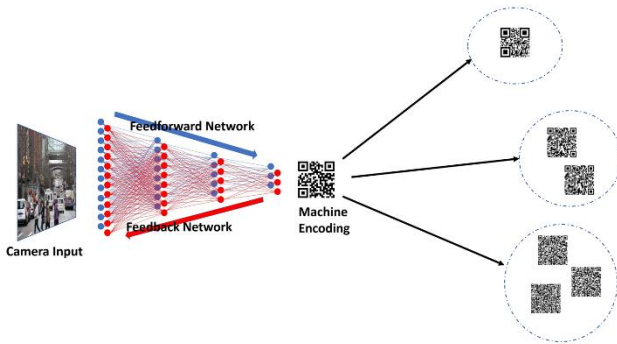
We need a more flexible connectome or network connection structure to train spiking neural networks. While DNNs only allow 'neurons' to connect to the next layer, connections in the visual cortex can go forward many layers, and even backwards, to form feedback loops. When two SNNs with complementary function and

opposite signal direction are organized into a feedback loop like this, Hebbian learning now helps train them to become an autoencoder, that is able to encode spatial-temporal inputs such as video, sound or other sensors, and reduce them to a compact machine representation and then decode that representation into the original input and together provide feedback to train this process.

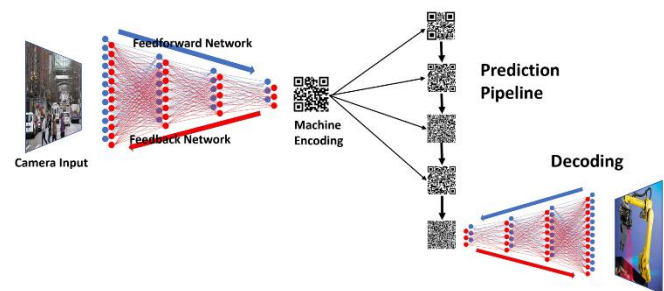


Notably, these compact representations will be clustered and organized in a multidimensional space where notions of distance correspond to our intuitive human understanding of similarity in the inputs. This brings the previously impossible task of clustering high-dimensional, high-memory video or speech down to the far simpler task of clustering the compact representations and using that clustering to do analysis or recommendation. Autoencoding of video and speech have intermediate commercial applications (in addition to being a step towards AGI). We can also put these encoded representations into other traditional ML algorithms like predictor pipelines and deep reinforcement learning, using their compact state to make predictions, or as states to decide on control outputs for a system.

Clustering – automatically grouping things



Sequence Prediction –What comes next?



Genetic algorithms are used to find the optimal feedback and autoencoder connectome design. Genetic Algorithms are made tractable for large spiking neural networks by representing these networks with a compact genome that is crossbred and mutated, then expanded through a deterministic and smoothly interpolating process to produce the full network connectomes to train and evaluate. Much like our biological genome that encodes the structure for billions of neurons and trillions of connections in only thousands of genes, our artificial genome specifies the layers and connection schemes between all the layers in the network in only a few KB of parameters. Optimal choices of parameterization and the process for expanding it to a connectome will encourage speedy convergence to global optima by directing the search toward useful structures while

maintaining sufficient diversity. Previous methods crossbred and adjusted synaptic weights directly, limiting genetic algorithms to very small networks because the parameter space (of all synaptic weights) was too large to search for larger networks.

Using these Bidirectional Interleaved Complementary Hierarchical Neural Networks (BICHNNs), constructed by our compact genome to full connectome expansions, we can efficiently perform genetic algorithms to specialize them into being optimal visual, speech, sensory, and even motion control cortices. Another novel behavior exhibited by these loops is that when properly set up and trained, when all inputs are turned off, they still hold internal state and continue to operate, meaning they have memory and logic. They don't yet dream of electric sheep, but this connectome structure can be evolved to do cognition or planning and give us a frontal cortex capable of complex decision making. In this manner, we can evolve most of the components we need to make an actual functional brain.

With that, we can assemble all the components we made into an artificial brain that would be functionally much more like a human brain, but we now architect it and evolve its macrostructure to best perform with the SNN technology, tools and processes used. An airplane does not flap feathered wings to fly, and actually works better with smooth aluminum skin and propellers. Similarly, an AI brain does not need all the exact characteristics of a human brain, just the right ones to do the job we can use artificial evolution to choose what to keep, what to substitute, and what to eliminate to build the artificial brain that optimally does the tasks we set it to.

Now how do we train this AI brain (with its collection of autoencoder cortices and frontal cortex decision making) to be human? We cannot transfer or copy a human consciousness from a biological brain to a synthetic one, as they will always be utterly incompatible, despite our best efforts, but we don't need to transfer a person's mind to our AI brain; we just need it to act the same as (or mimic) that person. We apply training data from performance capture of a specific human, including speech, textual correspondence and even body and facial motion capture to our AI brain to make it see, hear, talk, act and move a 3D body (or robot) like that person, becoming a digital mimic of them. Then we can evolve and scale these brains within their 'bodies', using their senses and outputs to interact with the user, scoring them during interactions and evolving them so they are better capable of speaking with us fluently and learning by observation, experience and practice, just like us.

We will probably not end up with an AGI like this, but we will have human mimic AIs that (when we add traditional computational, database, and deep learning capabilities for a specific job), will be narrow AIs, adept at a sufficient variety of localized tasks to function as a super-human AI employee doing a specific job. This is an important intermediate step to creating something with human-level or superhuman general intelligence and also has obvious immediate commercial value, putting these vocational AIs to work in customer service, information jobs, and even as AI assistants to high paying jobs like doctors, attorneys, financial analysts, and administrators.

Now how do we get from here to AGI? At a high level, we are now able to train these narrow AIs to each specialize at the tasks relevant to a specific job, and we make different versions of these AIs that work at thousands of different jobs from sales clerk, to desk agent, to concierge, to even doctors and lawyers assistants, and they learn further job and interpersonal communication skills when deployed. With this accomplished, we now have a large number of different narrow slices of AI that can come together like a pie. When networked and interconnected, this will create a very large and capable AGI precursor that can interact intuitively with humans

through speech, text and vision, like a person, doing so with many different people simultaneously, and do many of the non-physical tasks and jobs that humans do (better than we can do them). It may not yet be a complete AGI, but it will provide a scaffold on which to add further perceptual, cognitive, planning, interaction functionality, and vocational capabilities that were missing from our initial ensemble of narrow AIs. Once we have a scaffold of an AGI and a list of missing pieces, the job of creating a whole AGI gets much easier.

Now how do we now train and evolve an AGI that is getting filled in, and growing from near-human to superhuman in the training process? We simply continue to take input from the millions of people interacting with the different vocational AIs that it is now comprised of. As well, our realistic 3D animated mimic AIs could participate as external trainers, running in as many instances as we want, at superhuman speeds, to speed the training process up.

We would have users provide feedback (explicitly by the user giving each interaction a rating, or implicitly, measured by facial cues and body language) after every encounter to score the AGI they interacted with. Multiple AGIs would be active simultaneously, each serving a subset of the users. Periodically there would be a culling, and genes from the top 10% of the AGIs would be crossbred, and these new, improved instances trained and deployed, so that they are constantly evolving and improving. Each time they train, their networks are improved, and the accumulated set of global training data gets larger and richer, so they get better by both.

It will take a lot of training, scaling and evolution with some very sophisticated datasets and well-chosen selection criteria to first create these sensory, motor and cognitive cortices and to assemble them into artificial brains. It will take a massive enterprise to tailor thousands of them to specific jobs and deploy them commercially, then a global network of supercomputers, each about 5x the capacity of the ORNL Summit GPU supercomputer to run the final AGI candidates on, but all of this is quite feasible with current hardware technology roadmaps within 5-8 years. Adding neuromorphic chips that can run SNNs natively greatly reduces the computing hardware size and cost and makes deploying such advanced artificial intelligences much more economical and practical.

AGI is well within our grasp, and several companies, including [ORBAI](#) have announced their [intentions to pursue AGI](#). It is a worthy goal, as developing an AGI and scaling it beyond human capability, augmented with all our best technology in computation, information sciences, and access to the vast realms of data on the internet and interacting with billions of people every day will produce a very powerful, tool unprecedented in human history. This will enable us to make amazing discoveries in science, medicine, psychology, social sciences, and other fields. It would allow us vastly improve how we plan, distribute resources and administer our cities, countries, and our planet, to bring universal global nutrition, healthcare and education, to bring scrutiny to injustice, and to perhaps level the playing field for the rest of the world, righting many wrongs we have historically been unable to.

Brent Oster
CEO, ORBAI