

A Cell Is a Neural Network

Reading biology as deep learning

Rafael M. Ehlers · May 26, 2026

rafaelmehlers.com.br/essays/a-cell-is-a-neural-network

The May 27th issue of the [Latent Space newsletter](#) ran an analogy I keep coming back to. It started with the familiar idea that genes are like computer programs, then went a step further: the nucleus is storage, the ribosome is a JIT compiler, proteins are processes, signalling pathways are workflows. It's a clean mapping, and it gets further than the lazy "DNA is just code" version.

But the more I sat with it, the more I thought the computer framing leaves the most interesting part on the table. Cells aren't really like von Neumann machines. They're stochastic. The same "program" produces wildly different outputs in different contexts. The data and the code keep blurring into each other. None of this is weird if you stop comparing cells to laptops and start comparing them to the thing we've actually spent the last decade building.

A cell is a neural network. Or rather, the whole body is.

Start with how a vaccine works.

You inject a weakened or fragmented pathogen. Your immune system has never seen it before, so it generates B cells with a wide variety of randomly-shuffled antibody shapes, selects the ones that bind, and keeps a memory of them. Next time the real version shows up, the system recognizes it and responds fast.

That's fine-tuning, structurally.

Your immune repertoire after general development is the base model. The vaccine is the training data. Selection for binding affinity is gradient descent on a loss function. The memory B cells are the updated weights. The next encounter with the pathogen is inference at deployment.

The failure modes map just as cleanly.

Allergies are overfitting. The system has learned that pollen, or peanuts, or cat dander is a threat. It isn't, but the model has fit a spurious correlation and now reacts strongly to harmless inputs. Antihistamines are output filters: they don't fix the model, they suppress the wrong reaction at runtime.

Autoimmune disease is a jailbreak. The system has lost the ability to distinguish self from non-self and starts attacking its own tissues. The safety check that says "don't activate against this" has degraded. Lupus, type 1 diabetes, rheumatoid arthritis: these are alignment failures in a learned classifier.

Immunodeficiency is a model with the safety training stripped out. Without functioning T cells (HIV is the famous case), the system stops distinguishing pathogen from harmless microbe, and

ordinary bacteria become lethal. Same pattern as removing refusal training from a language model: the discriminator collapses and everything passes through.

Once you see the immune system this way, the rest of the biology follows. DNA isn't really source code, it's the trained weights: compressed, opaque, redundant, only legible because we've spent decades doing what amounts to mechanistic interpretability on it. Gene expression is which features fire at a given moment. Cell types are different fine-tunes of the same base model. A liver cell and a neuron share weights, they just have different parts active. Cancer is what happens when enough mutations accumulate that the network starts optimizing the wrong objective.

What's striking is that the tools are converging. Protein design (RFdiffusion, AlphaProteo, ESM3) and model editing (activation steering, feature ablation, fine-tuning on specific behaviors) are doing the same kind of work: searching a high-dimensional design space for compositions of functional units that produce a specified behavior.

If that's right, the implication is bigger than the analogy. Biology was the original foundation model. Four billion years of pretraining on an environment, a vast learned vocabulary of composable parts, and a deployment surface that now numbers in the trillions of instances. Deep learning finally gave us a neural substrate we can train in less than a geological epoch.