

Case Studies

Study Guide

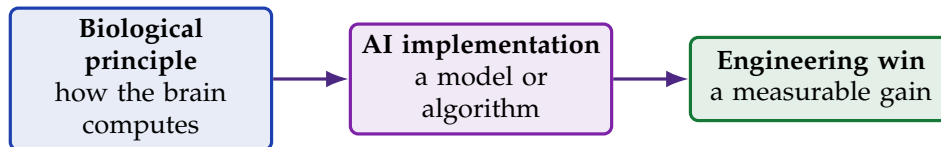
by Richard J. Young, Ph.D.

richard@deepneuro.ai

Chapter 23 of "The NeuroAI Handbook"

Four times a brain principle became a working algorithm: predictive coding, replay, attention, and fusion.

One-sentence thesis. NeuroAI's payoff is a repeatable recipe: name a principle for how the brain computes, build it into a model, and collect a measurable engineering win.



Every case in this chapter runs left to right along this pipeline. The lesson is the pipeline itself.

PART 1: Predicting the World

| The Principle: A Brain That Guesses

Start with a surprising claim: the brain is not a passive camera, it is a prediction machine. In the **predictive coding** framework, higher areas constantly send top-down guesses about what the lower sensory areas should be seeing, and the lower areas send back only the **prediction error**, the mismatch between guess and reality (Rao and Ballard, 1999). Everything already predicted is silence; only the surprise travels upward. It is a ruthlessly efficient way to perceive.

In symbols, each level transmits not its raw input but the residual:

$$\varepsilon = \text{input} - \text{prediction}$$

ANALOGY

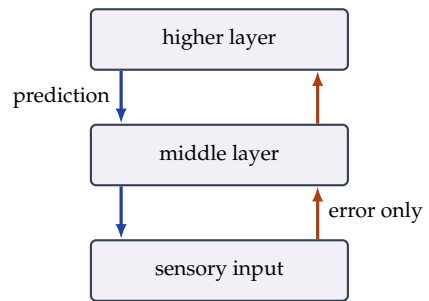
A good copy editor does not re-read you the whole manuscript, they hand back only the marked-up errors. Predictive coding runs perception that way: the page is assumed correct, and only the typos, the prediction errors, are worth sending up the chain.

| The Model: PredNet

PredNet (Lotter, Kreiman, and Cox, 2016) wires this into a recurrent convolutional network. Every layer tries to predict the activity of the layer below it; the top-down prediction is subtracted from the actual input, and only the resulting error is passed up to be explained.

| The Win

Predicting the next video frame, PredNet beat standard feedforward models, and the features it learned, with *no labels at all*, transferred to more robust object recognition, especially on noisy or ambiguous inputs. **Takeaway.** a model that actively predicts its own input, as the brain does, learns richer representations for free.



Predictive coding, and PredNet: guesses flow down (blue), only the surprises flow back up (amber).

■ The Principle: Learning During Rest

The brain does not learn only in the moment. The **hippocampus** re-plays the neural sequences of recent experience during sleep and quiet rest, a process thought to consolidate memories and hand them to the neocortex for long-term storage (Foster and Wilson, 2006). Crucially it does not replay everything equally, it favors the surprising and the important.

■ The Model: Prioritized Experience Replay

Reinforcement-learning agents already used **experience replay**, storing past transitions in a buffer and sampling them randomly to break the temporal correlations that destabilize training. DeepMind's **Prioritized Experience Replay** (Schaul et al., 2015) added the hippocampal twist: sample by *surprise*. The measure of surprise is the **TD-error**, how far the outcome fell from prediction, and high-error transitions are replayed more often.

In symbols, the surprise signal and the sampling rule are:

$$\delta = r + \gamma \max_a Q(s', a) - Q(s, a), \quad P(i) \propto |\delta_i|^\alpha$$

🔗 ANALOGY

Smart studying is not re-reading the whole textbook, it is drilling the flashcards you keep getting wrong. PER gives the agent that instinct: spend repetitions where the prediction failed, not where it already succeeds.

■ The Win

The tweak was small and the effect large: agents learned Atari games to superhuman level **faster** and on far **less data**, and reached higher final scores, purely by replaying the informative memories first.

The Principle: Global Context

The visual cortex is hierarchical, but it is also laced with long-range horizontal and feedback connections that let one part of a scene inform the reading of another. This global integration is the substrate of **visual attention**: selectively weighting what matters given the whole context, not just the local patch.

The Model: The Vision Transformer

For years vision meant CNNs, which see only local receptive fields and build up context slowly. The **Vision Transformer** (Dosovitskiy et al., 2020) threw that out. It slices an image into fixed patches, treats each patch like a word in a sentence, and feeds the sequence to a Transformer whose **self-attention** lets every patch weigh every other patch, near or far, from the very

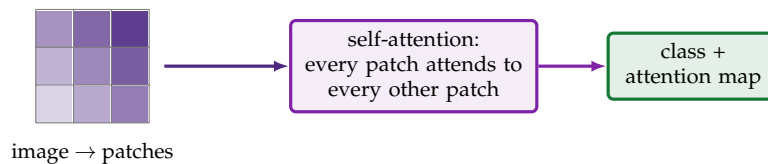
first layer (Vaswani et al., 2017).

Formally, attention scores every query against every key, then mixes the values:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) V$$

The Win

Trained on large enough data, ViTs **surpassed** state-of-the-art CNNs. They capture relationships between distant image regions, scale more gracefully as data and model grow, and their attention maps are directly **interpretable**, a literal picture of where the model looked.



The Vision Transformer: patches become tokens, and self-attention integrates the whole image from the first layer.

PART 4: Fusing the Senses

The Principle: Multisensory Integration

The brain builds one robust percept by pooling many senses. A region like the **superior temporal sulcus** responds to a face, a voice, and a name together, and that fusion yields a model of the world more reliable than any single channel. The whole is steadier than its parts.

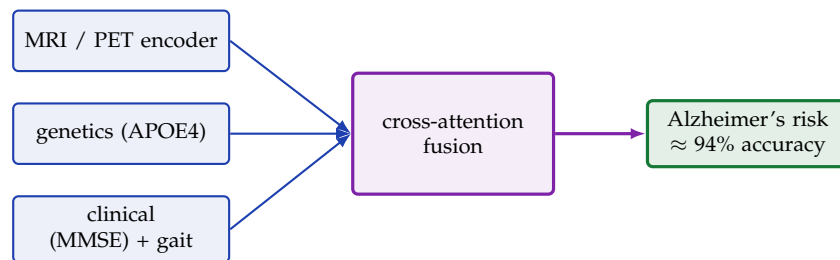
The Model: Multimodal Fusion

A patient, like a percept, is not one number. Inspired by the brain's integration, **multimodal models** for Alzheimer's prediction fuse very different data streams: **neuroimaging** (MRI and PET), **genetics** (risk

genes such as APOE4), **clinical** data (cognitive scores like the MMSE, demographics), and **behavioral** signals (speech, gait). A dedicated encoder reads each modality, and **cross-attention** learns how the streams relate before a joint prediction (Jack et al., 2018).

The Win

Fusing the evidence beat every single-modality model, reaching about **94%** accuracy in the chapter's example, flagged high-risk individuals **years** before clear symptoms, and surfaced **personalized** biomarker combinations, the mix most predictive for a given patient.



Multimodal fusion: one encoder per data type, cross-attention to relate them, a prediction stronger than any modality alone.

Brain principle	AI implementation	Signal used	Engineering win
Predictive coding	PredNet	Prediction error	Robust, label-free vision
Hippocampal replay	Prioritized replay (PER)	TD-error (surprise)	Far better sample efficiency
Global integration	Vision Transformer	Self-attention	Scales past CNNs
Multisensory fusion	Multimodal fusion	Cross-attention	Earlier Alzheimer's prediction

Quick Self-Test

1. In predictive coding, what does a lower area send up to a higher one, and why is that efficient?
2. What signal does Prioritized Experience Replay use to decide which memories to replay more, and what brain phenomenon inspired it?
3. How does a Vision Transformer turn an image into something a Transformer can attend over, and what can it do that a CNN struggles with?
4. Name the four data streams fused in the Alzheimer's model, and state the payoff of fusing them.
5. State the three-stage NeuroAI recipe that all four cases share.

If You Remember One Thing

Each case is the same pipeline: a principle for how the brain computes becomes an algorithm, and the algorithm delivers a measurable win. Predict your input (PredNet), replay your surprises (PER), attend to everything at once (ViT), fuse your senses (multimodal). The brain is not just inspiration, it is a working design document.

Richard J. Young, Ph.D. · richard@deepneuro.ai · Senior AI Research Scientist, UnitedHealth Group · UNLV Lee Business School · ORCID 0000-0002-1109-7552