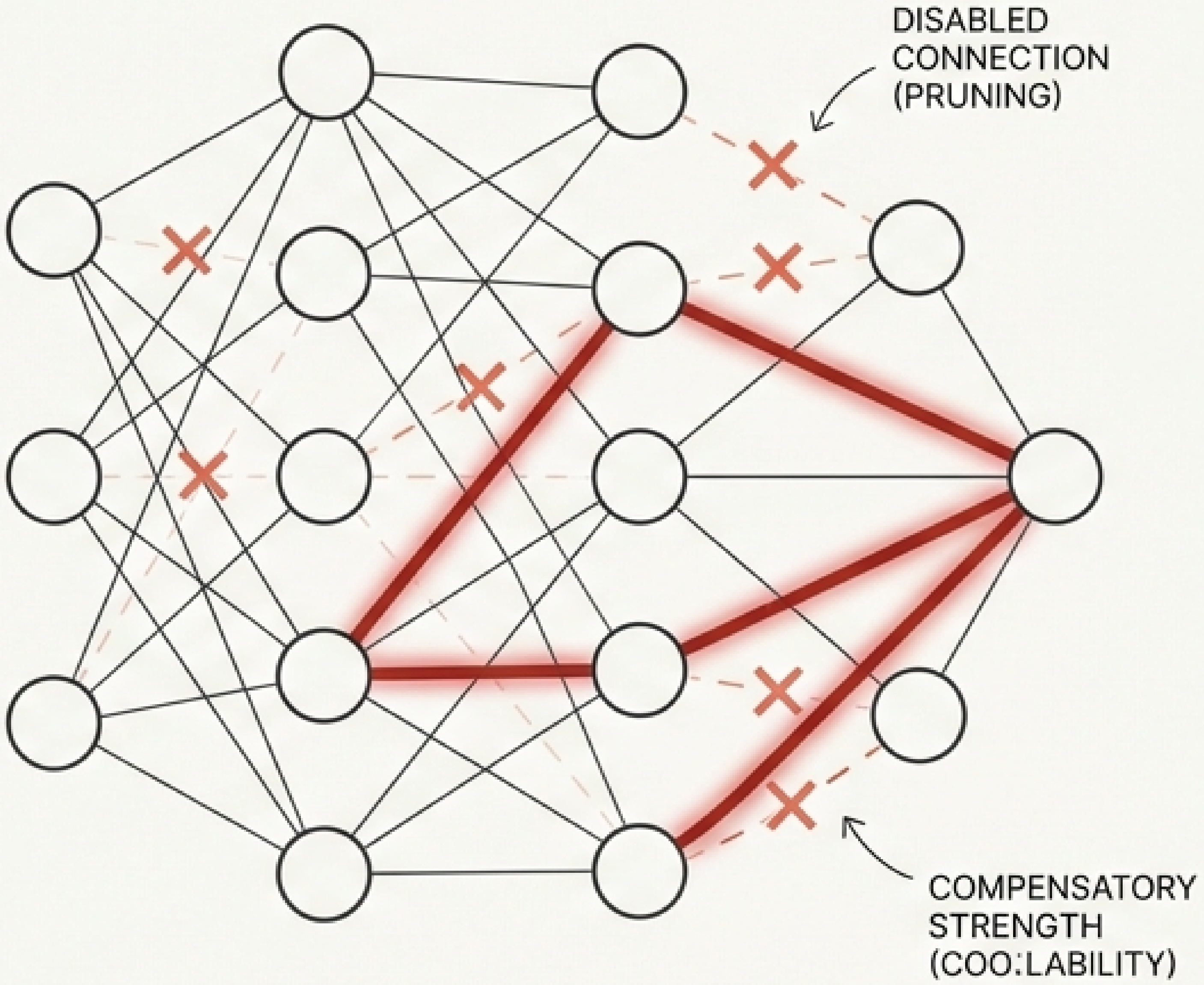
Neural Networks with Disabilities

How Training AI with Limitations Creates Unprecedented Resilience



Based on the research of V. Terziyan and O. Kaikova, *Neural Computation* 34 (2022).

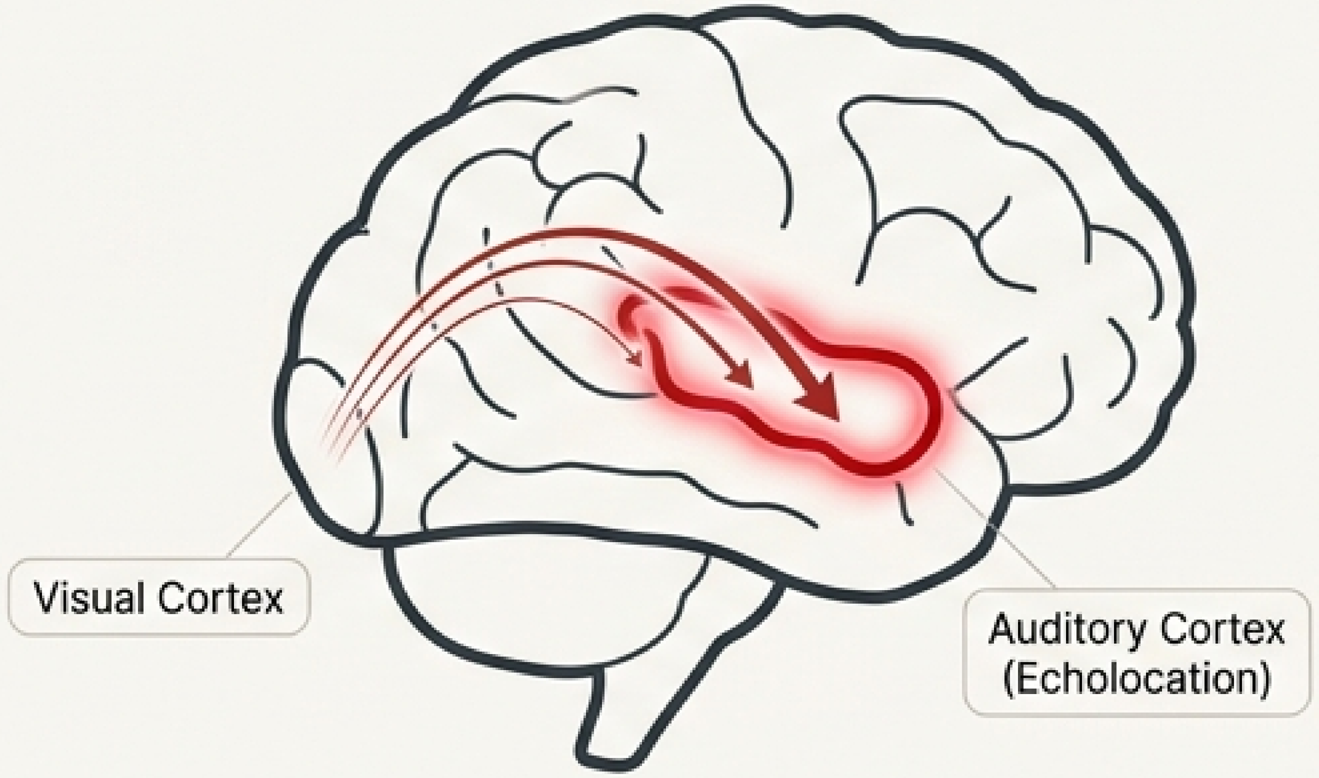
The Coolability Paradox: How can less become more?

The Human Analogy

Disability often co-occurs with enhanced abilities, or "**coolabilities**." These are the strengths that emerge in response to a disabling condition.

Example 1: Blind individuals can **reassign** the brain's visual cortex to develop superior echolocation.

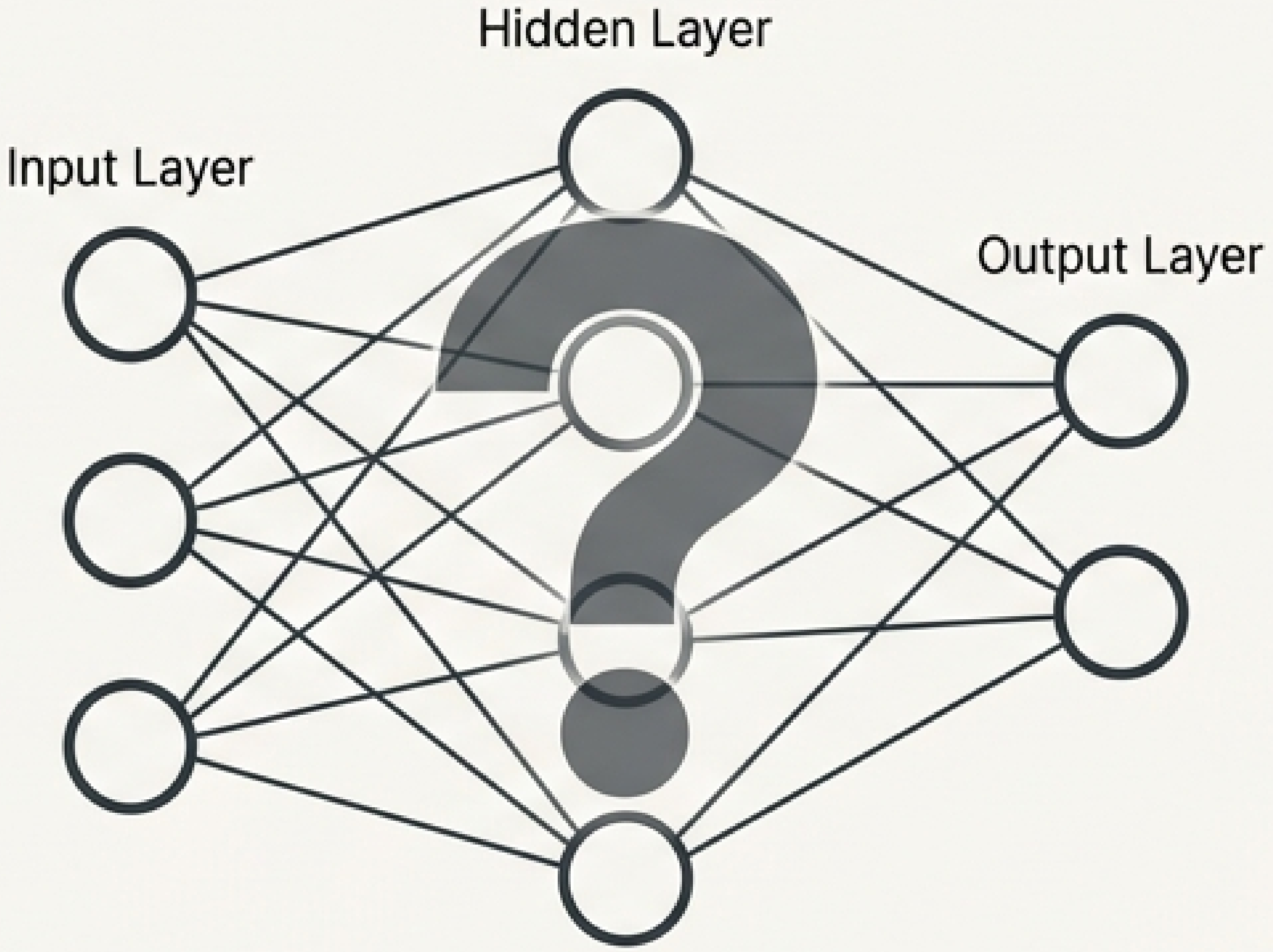
Example 2: Athletes train blindfolded to **amplify other senses** and improve proprioception, leading to greater mastery.



The Central Question for AI

If evolution and training can **induce coolabilities** in humans, can we intentionally induce them in AI?

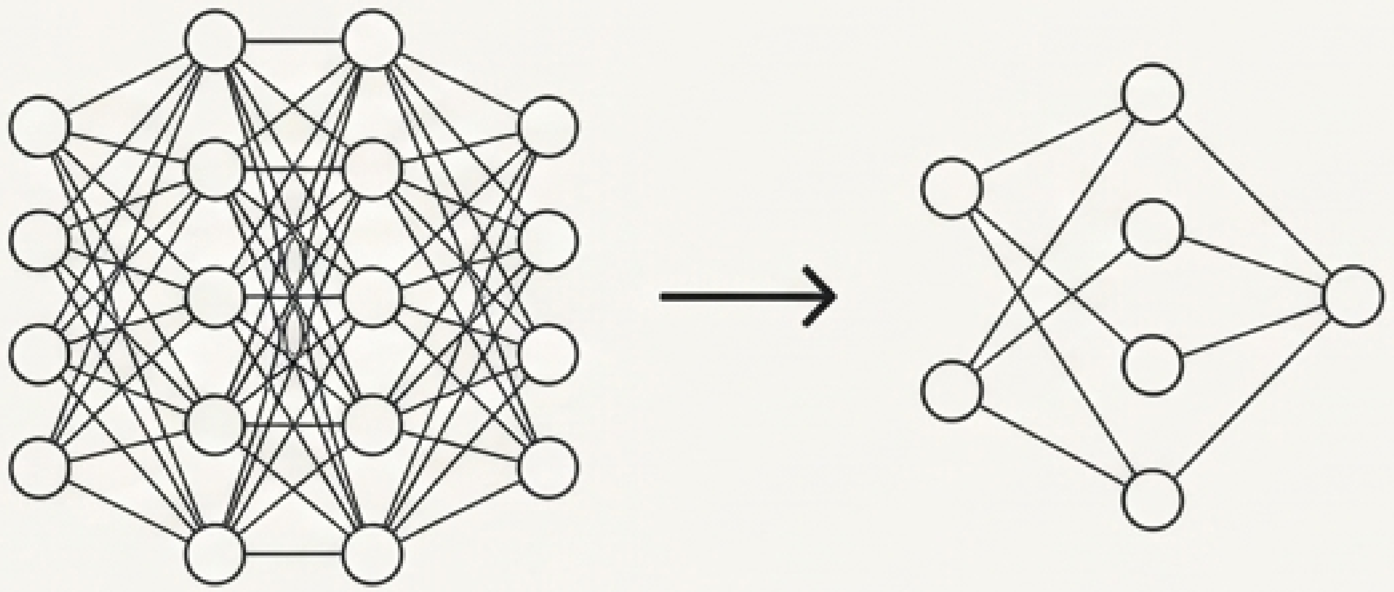
Hypothesis: **Intentionally disabling** AI models during training will lead to **superior performance** and **resilience**, especially under adverse conditions.



A Paradigm Shift in Pruning: From Efficiency to Resilience

The Old Paradigm (Pruning for Efficiency)

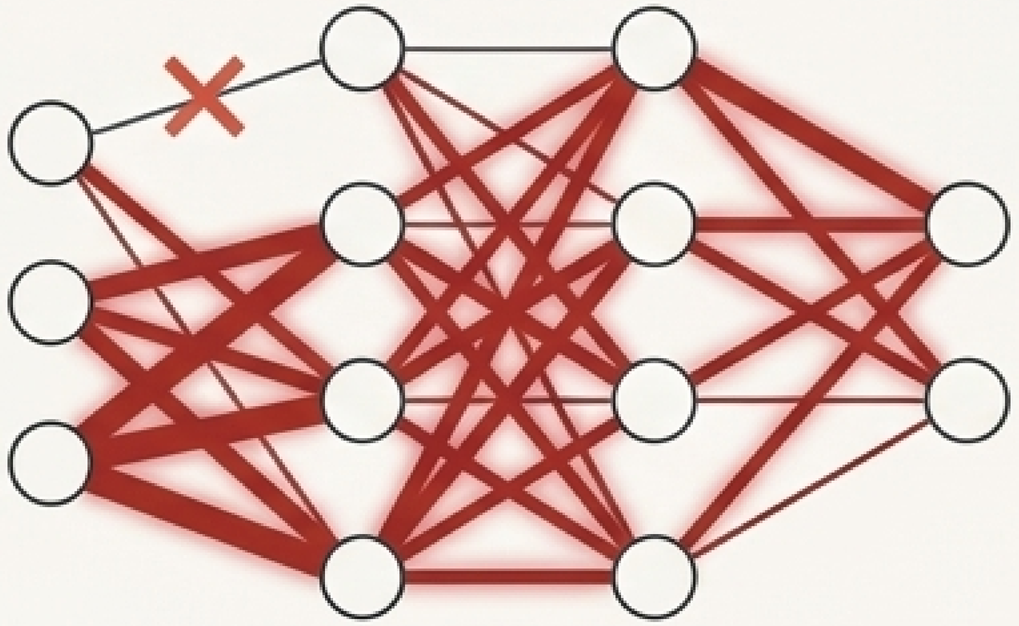
- **Techniques:** Dropout, traditional pruning.
- **Mechanism:** Temporarily or permanently remove neurons or connections to induce sparsity.
- **Primary Goal:** Combat overfitting and improve generalization by removing redundant information. The focus is on creating smaller, more efficient models.



Before After

The New Paradigm (Pruning for Resilience)

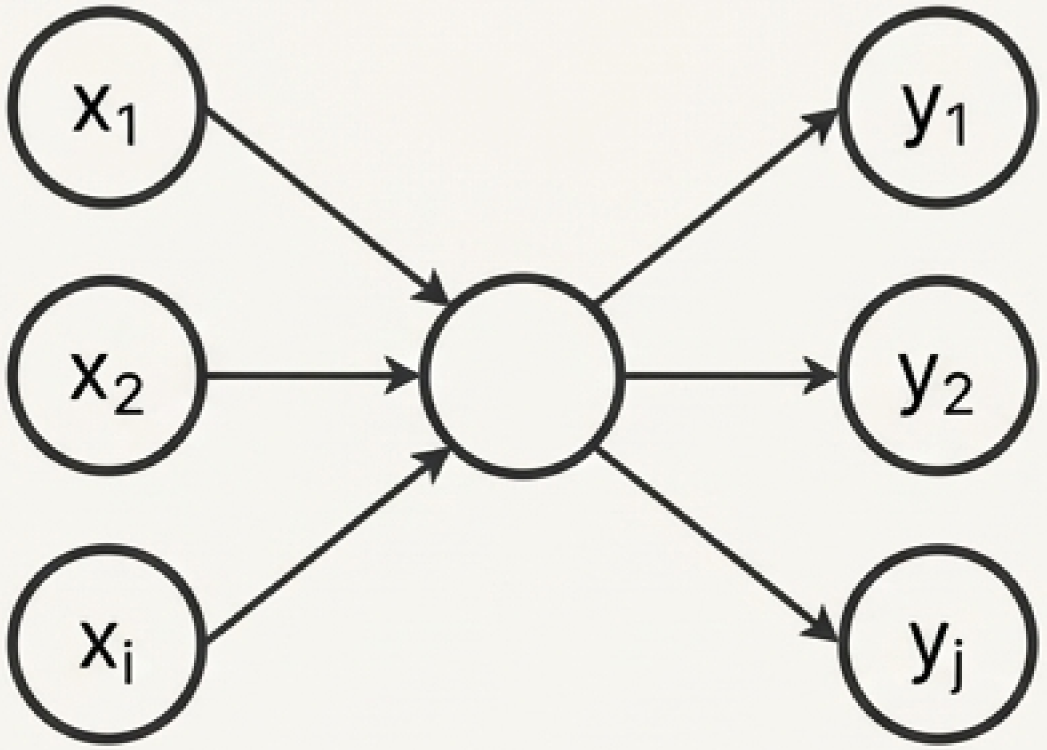
- **Techniques:** Complementary Artificial Intelligence (CAI).
- **Mechanism:** Introduce structured, *meaningful* disabilities by simulating specific functional or sensory losses.
- **Primary Goal:** Pre-train the network for adversity, creating ‘coolabilities’ (compensatory skills) that lead to systemic robustness. The focus is on creating stronger, more resilient models.



The Atomic Unit: Defining Disability for a Neuron

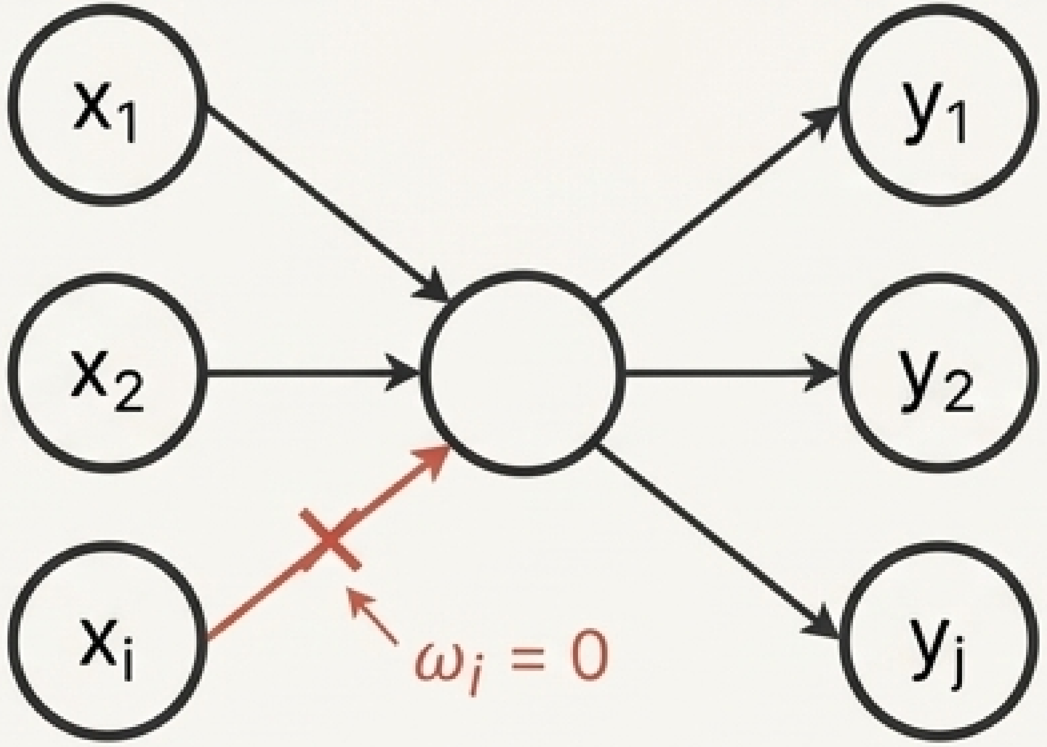
The Healthy Neuron

Fully connected. Receives all inputs and contributes to all outputs. Information flows freely forward and backward during training.



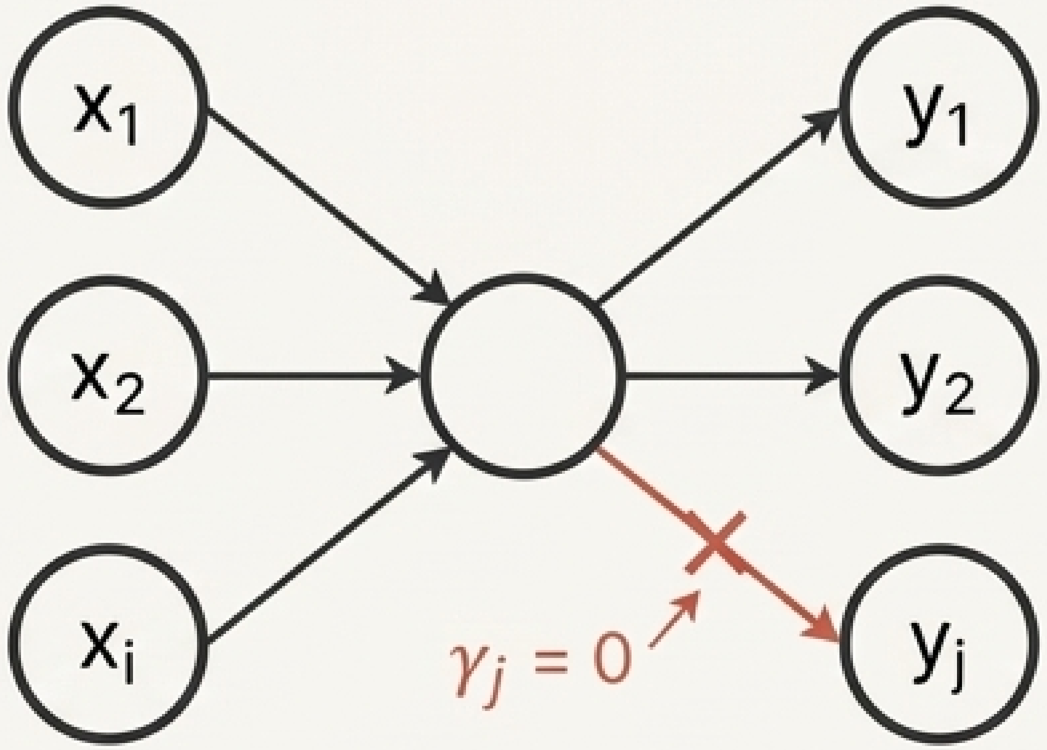
Sensory Disability

An input connection is permanently pruned ($\omega_i = 0$). The neuron is forced to learn its task ***without*** that specific piece of information.



Functional Disability

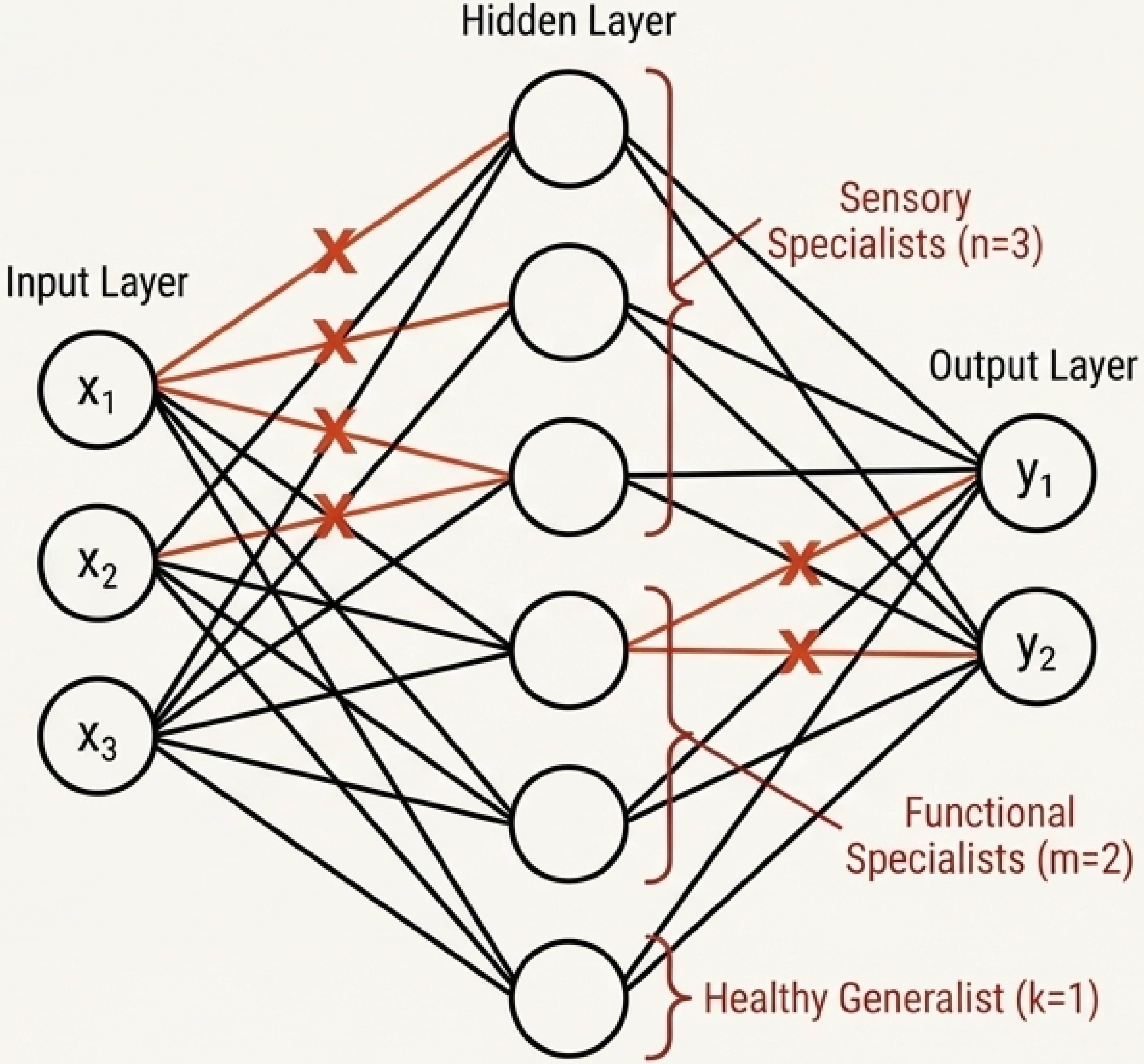
An output connection is permanently pruned ($\gamma_j = 0$). The neuron learns to achieve its goal ***without*** influencing that specific output or action.



A Shallow Network Becomes a Team of Specialists

The Concept: Instead of a hidden layer of identical "generalists," we build a layer of diverse specialists and generalists.

- "n" neurons, each with a unique sensory disability (one for each input "x").
- "m" neurons, each with a unique functional disability (one for each output "y").
- "k" healthy "generalist" neurons to integrate information.



The Key Insight: This disability-driven network is not just a collection of weaker units. It forms a **more robust system** that often outperforms a healthy network with the ***exact* same total number of connections**.

This is the coolability paradox in practice.

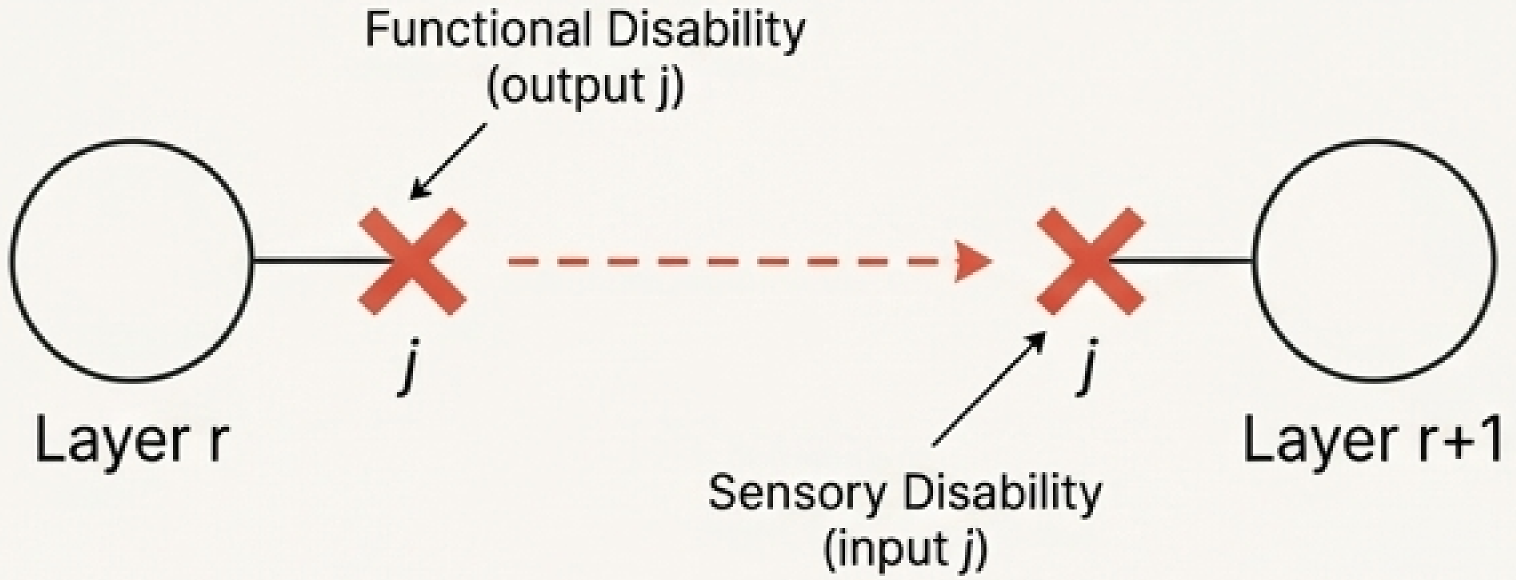
Propagating Disability Through Deep Architectures

The Challenge

How do you connect a layer of functionally disabled neurons to the next layer?

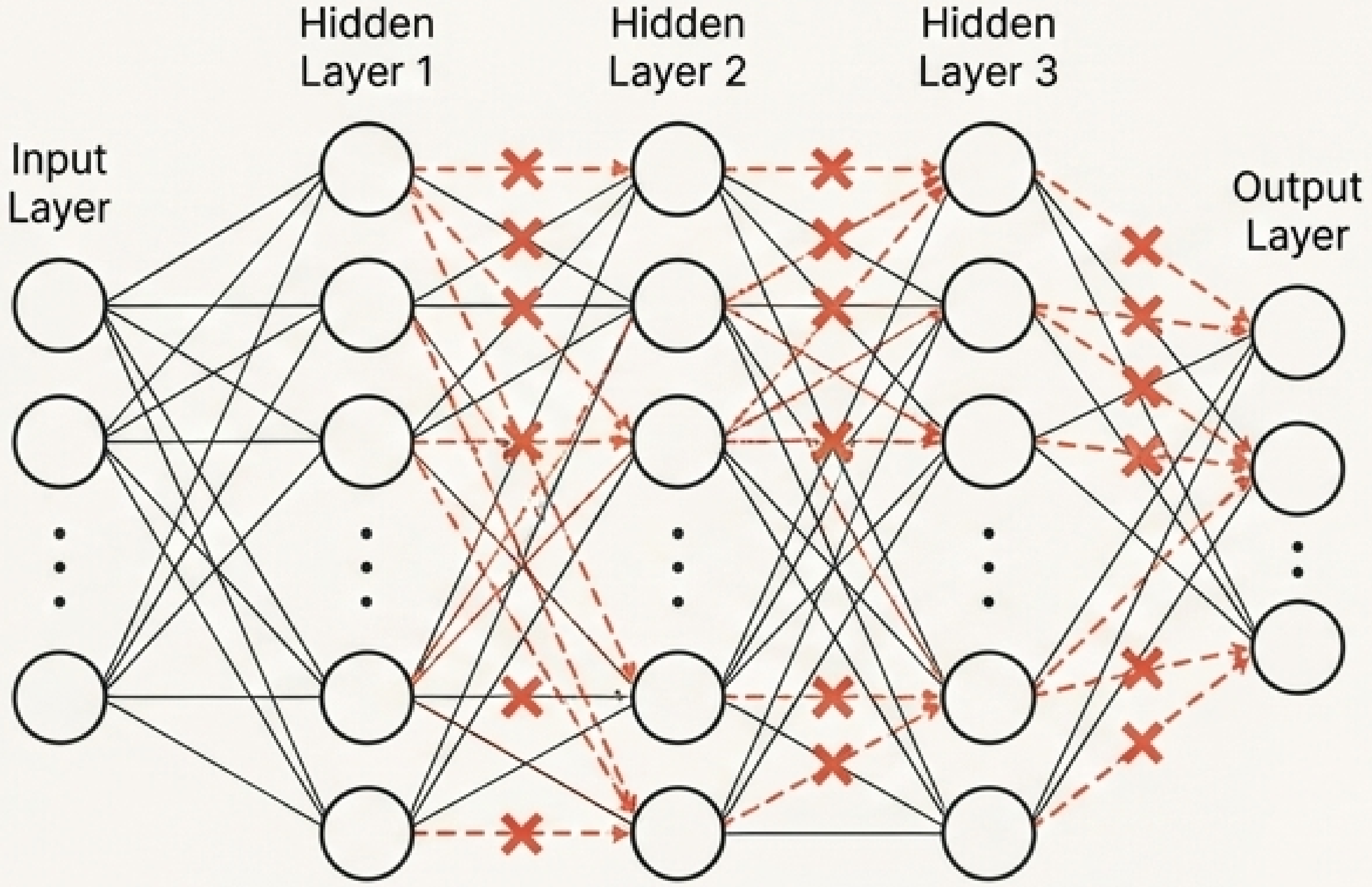
The Solution: The “Divorced Couple” Constraint

If a neuron in layer r has a functional disability affecting output j, then the corresponding neuron in the next layerr+1 must have a sensory disability for input j.



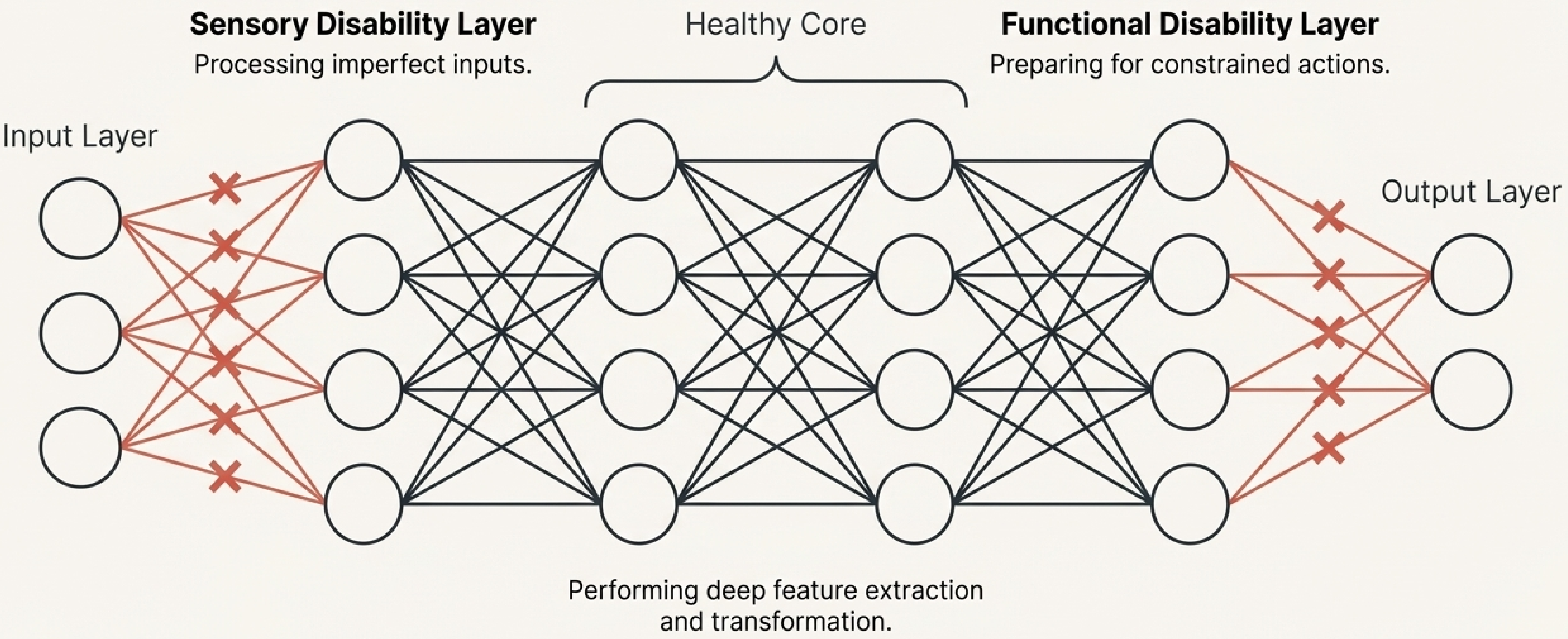
Deep Architecture

This rule ensures the pruned connection is consistent across layers, creating a deep web of compensatory learning pathways.



A Hybrid Architecture: Simulating Disability at the Cognitive Interface

A more realistic model where disabilities are isolated at the system's "edges," leaving a healthy core for complex processing.



Extending the Concept to Recurrent Networks: Memory Disability

- **The Concept**

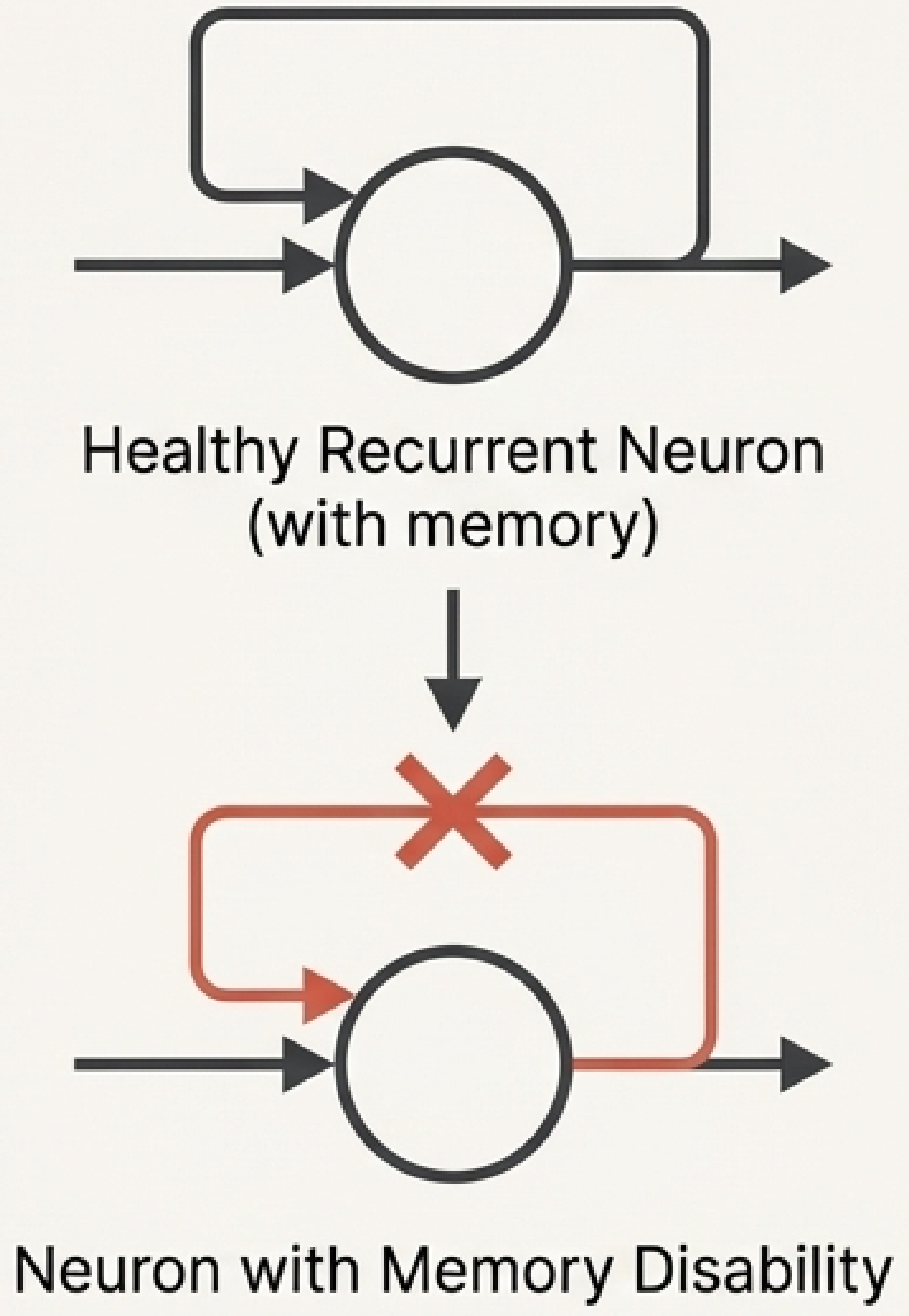
  In RNNs, a neuron's connection to its own previous state constitutes its "memory." We can treat this connection like any other.

- **Memory Disability**

  Pruning the recurrent loop (`neuron -> self`) of a specific neuron.

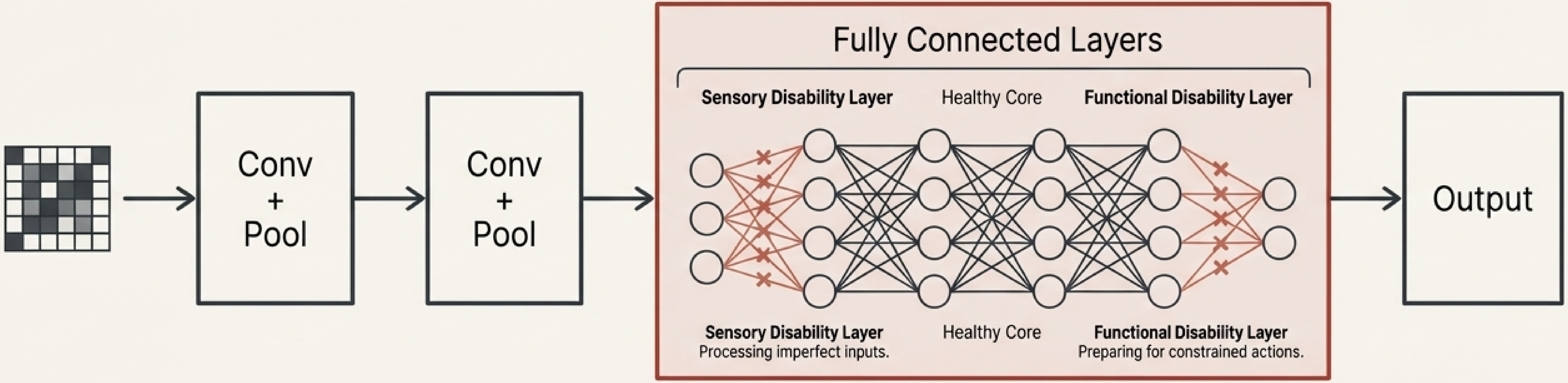
- **The Effect**

  This forces the network to find alternative pathways to retain information over time, potentially making it more robust to temporal noise or interruptions. This principle applies to both simple RNNs and more advanced architectures like LSTMs.



Hardening Computer Vision Models with Disabled Layers

- **The Application**: Standard Convolutional Neural Networks (CNNs) use convolutional layers for feature extraction and fully connected (FC) layers for final classification.
- **The Method**: The disability-driven architectures we've discussed (shallow, deep, hybrid) are embedded into the FC part of astandard CNN. The convolutional layers remain unchanged.

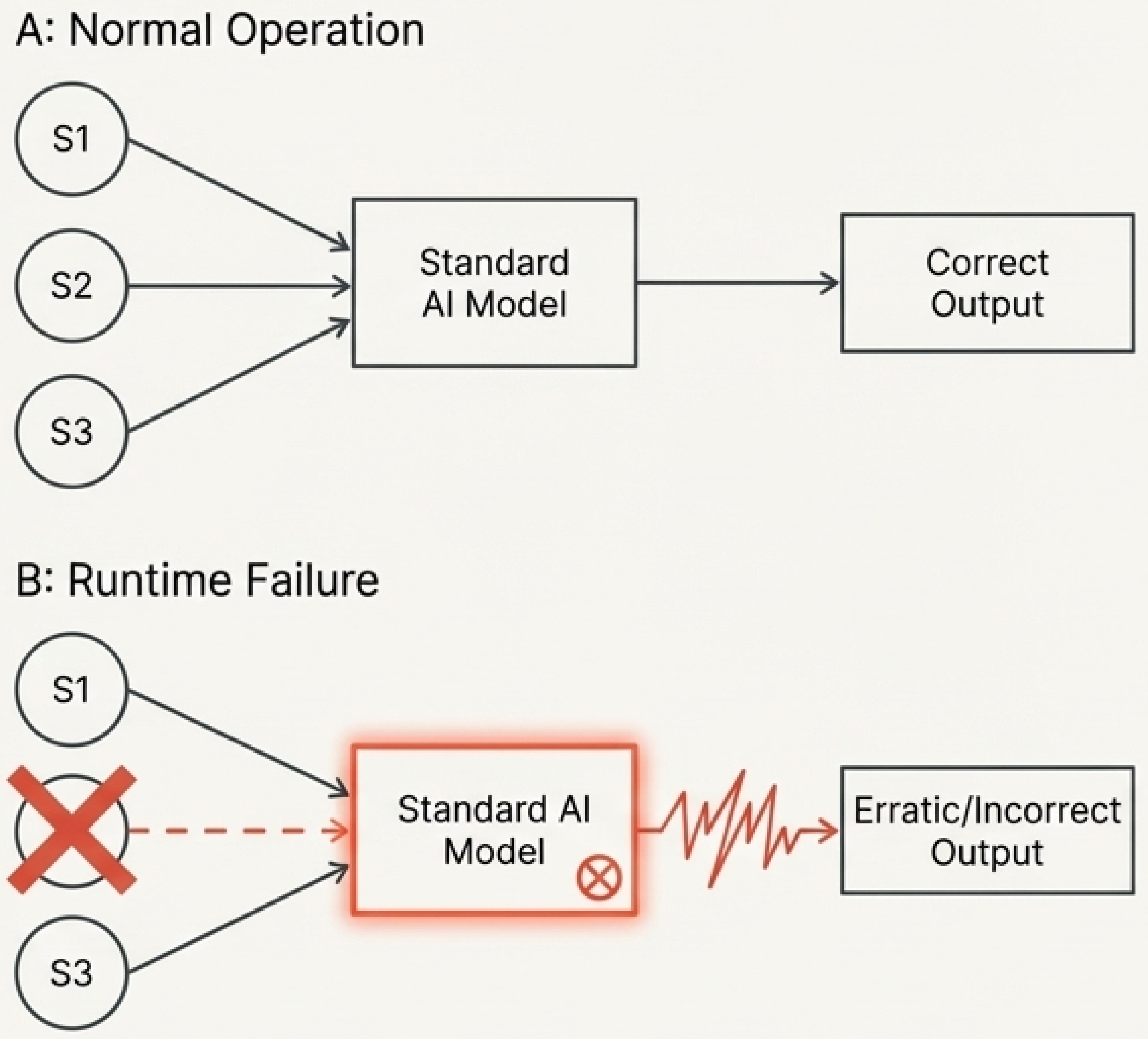


- **The Result**: A CNN that is more resilient to adversarial attacks or corrupted image data.
- **Evidence from NATO Project**: In a NATO SPS cyber-defense project, applying these principles resulted in a **2% to 10% classification precision improvement** against adversarial attacks.

The Ultimate Goal: Moving from Reactive to Proactive AI

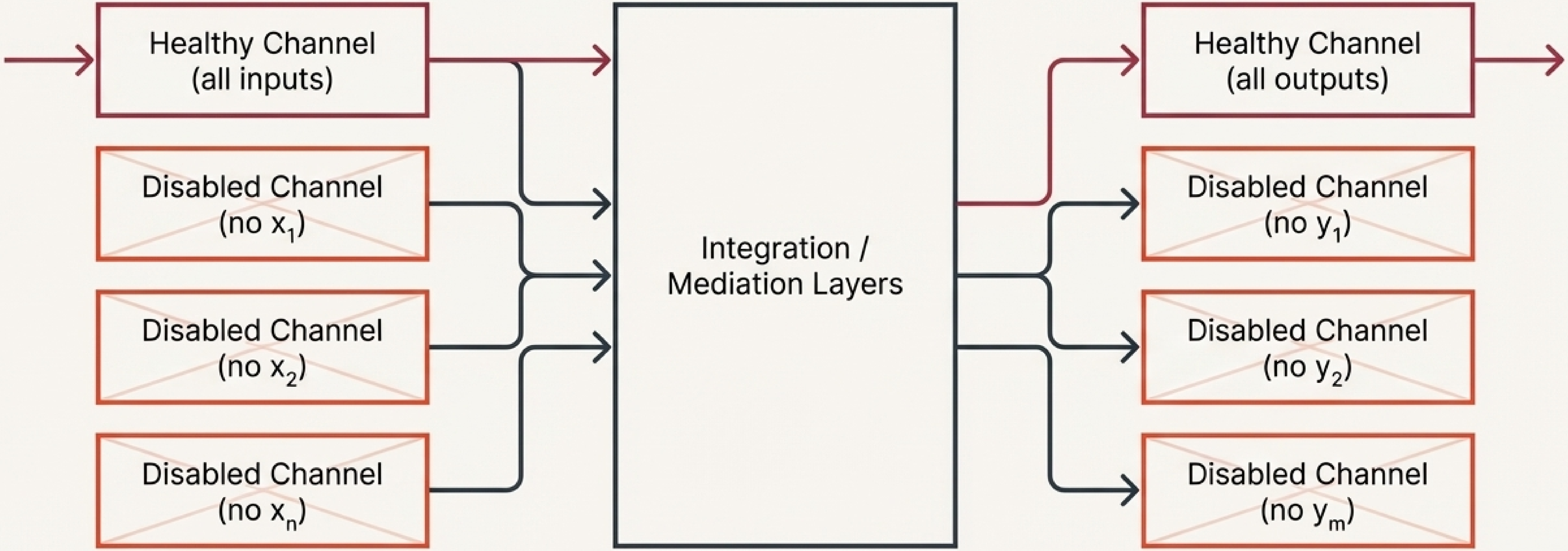
- **The Problem with Standard Models:** An optimized, healthy model is brittle. It's trained on the assumption that all sensors and actuators are always available. At runtime, if a single sensor (input) or actuator (output) fails, its performance can collapse because it has never encountered that specific scenario.

- **The Solution:** A system explicitly pre-trained for every likely point of failure. It doesn't react to failure; it anticipates it.



The Sustainable Multichannel Network: A Blueprint for Resilient AI

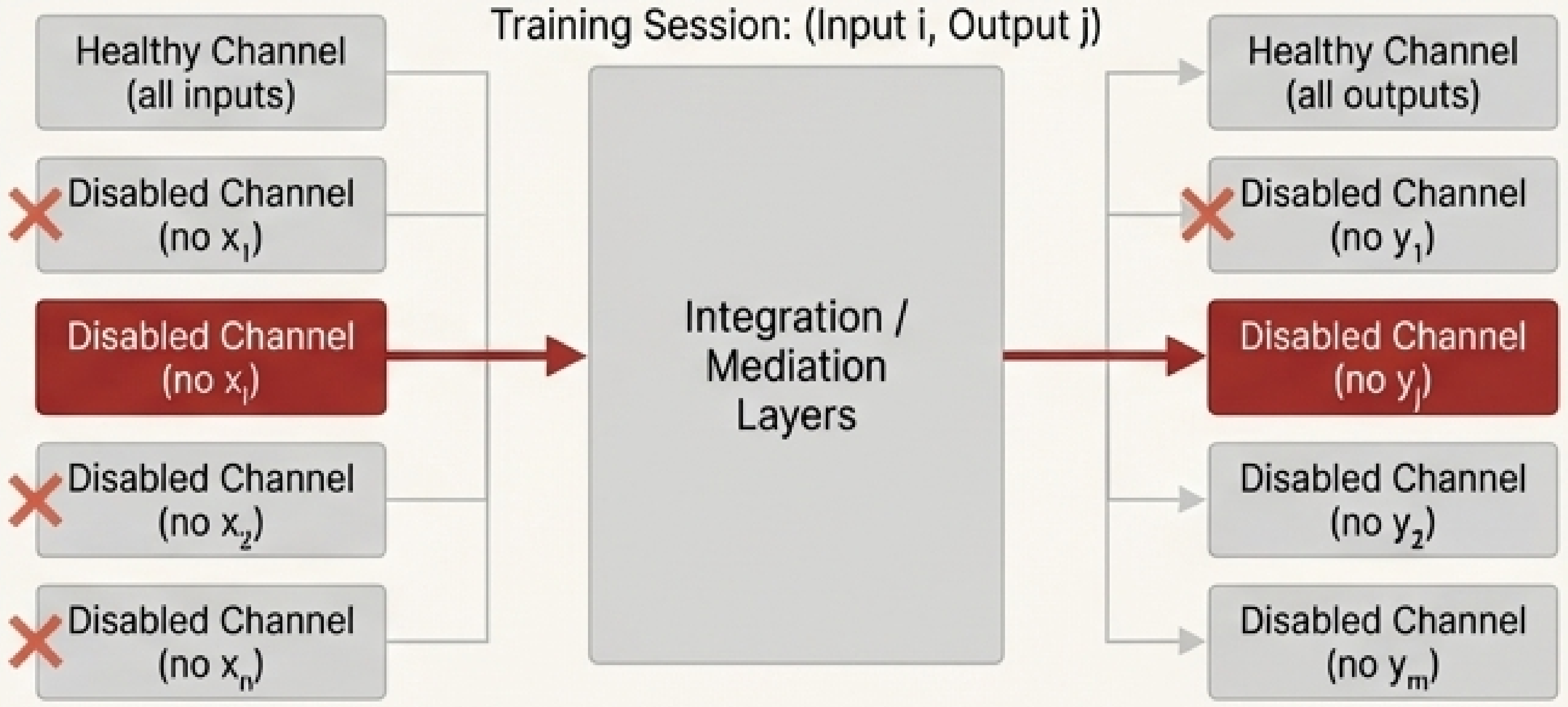
An architecture with parallel, isolated "channels," where each channel is a specialist sub-network pre-trained for a specific failure scenario.



Training for Every Contingency, Operating with Instant Adaptation

Training Phase

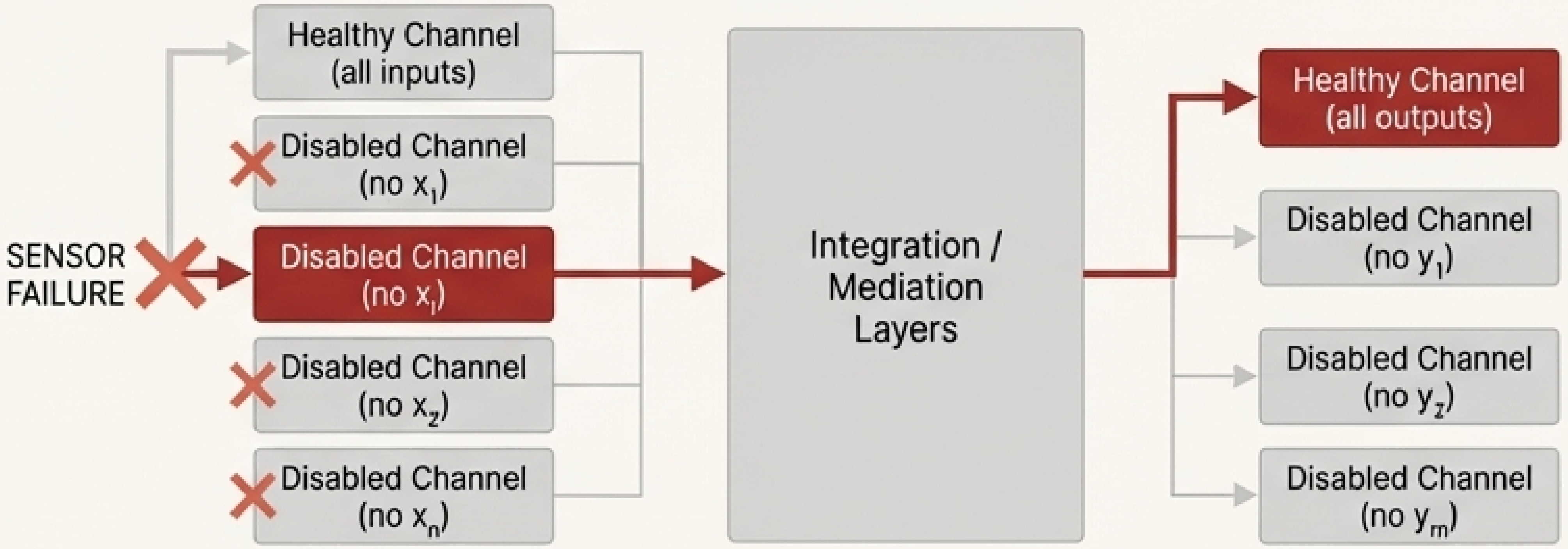
- The network is trained in `(n+1) x (m+1) separate sessions.
- In each session, only one pair of (input channel, output channel) is active and trained. All other weights in the network are frozen.
- This creates a comprehensive library of specialized sub-networks, each one an expert for a specific disability combination.



Operation Phase (Runtime)

Healthy State: The system uses the default (`healthy-in`, `healthy-out`) channel.

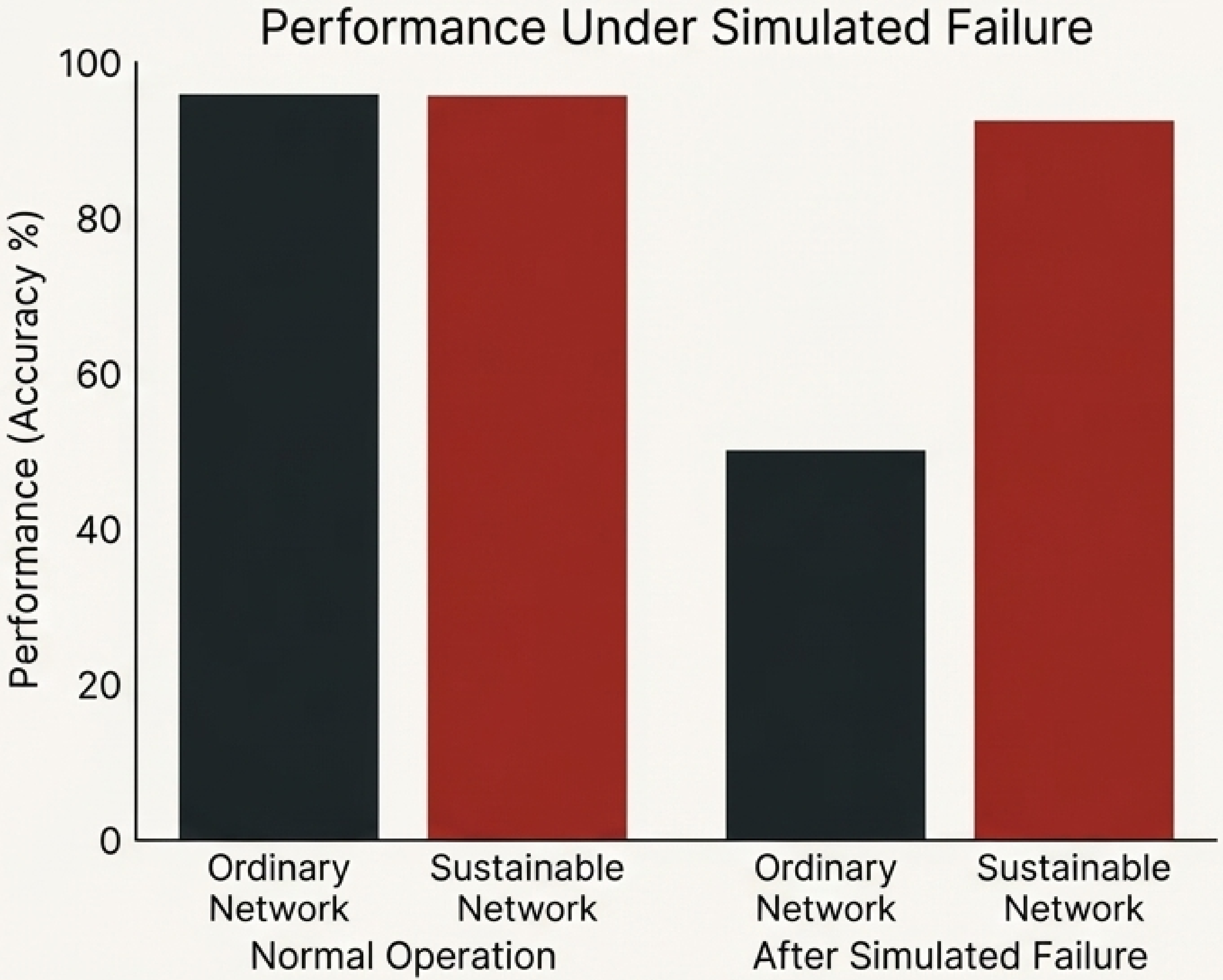
Failure State: If sensor `i` fails, the system instantly switches to routing data through the pre-trained (`disabled-input-i`) channel. The 'coolability' is already baked in. No retraining is required.



Measured Impact: Proven Resilience in Adversarial Conditions

- **Finding 1: Synthetic Data**
  In experiments with sudden input or output failures at runtime, the sustainable architecture **outperforms an ordinary network by up to 20%** in some cases.

- **Finding 2: Real-World Application (NATO Project)**
  Applying these principles to CNNs for cyber-defense resulted in a **2-10% classification precision improvement** against adversarial attacks.



Disability – Disability is Not a Bug, It's a Feature

By treating **structured disability as a training tool rather than a random flaw**, we create AI systems that are **inherently more robust and resilient**.

Complementary AI (CAI): A new class of techniques designed to pre-train AI for compensation, making it prepared for adversarial environments and real-world damage before they occur.

The Vision: Envitas the proactive approach to resilience to caphalse trustworthy AI in mission-critical domains.

- **Autonomous Systems** (e.g., self-driving vehicles)

- **Critical Infrastructure Management**

- **Cyber-Defense**

Further Reading: "Neural Networks With Disabilities: An Introduction to Complementary Artificial Intelligence" – V. Terziyan and O. Kaikova, *Neural Computation* 34, 255–290 (2022).