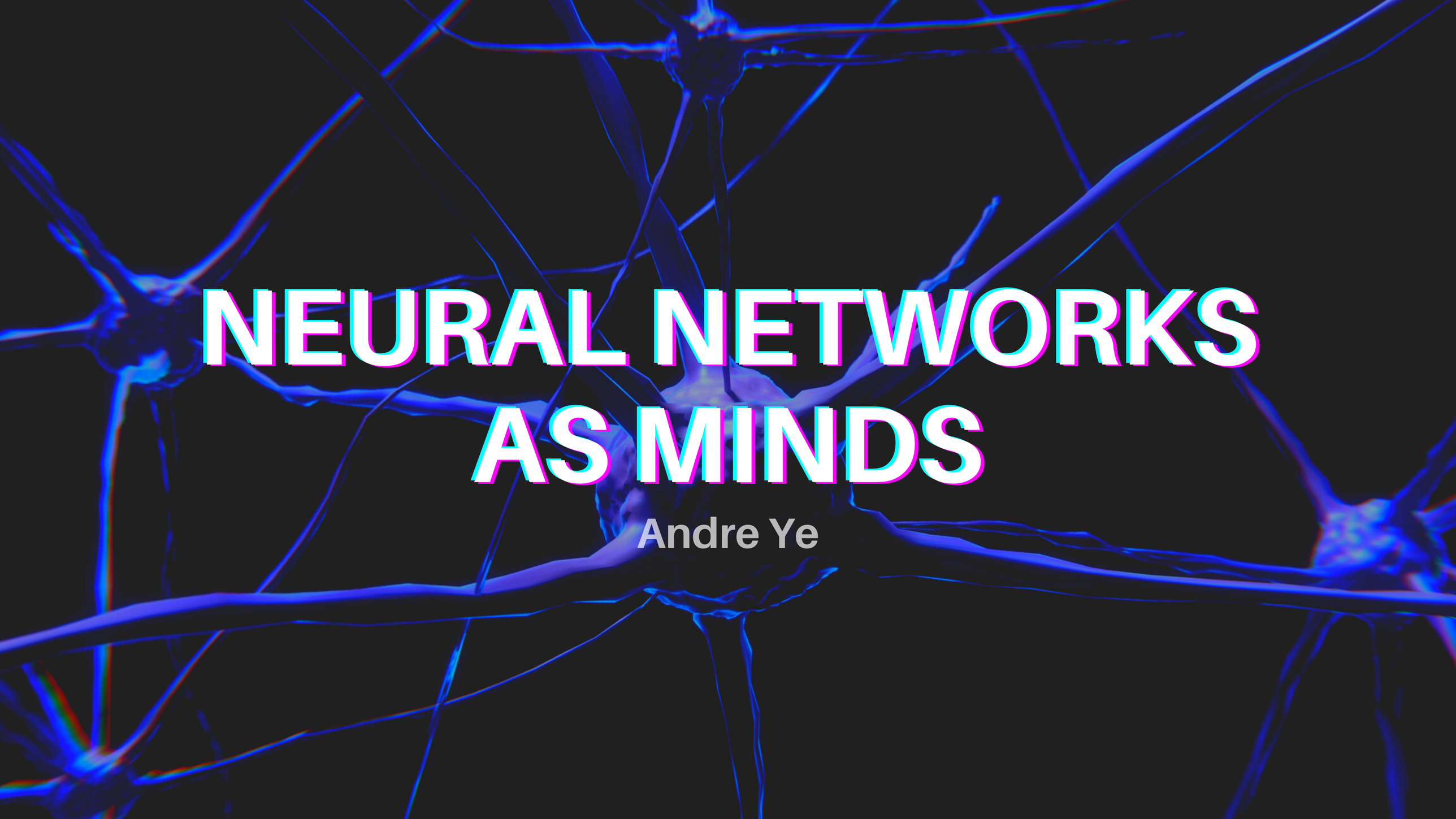

NEURAL NETWORKS
AS MINDS
Andre Ye

Agenda

To explore loose parallels between 9 deep learning research fields and traits of 'intelligence', starting from simple ideas to more complex ones

Objective

To inspire cross-disciplinary interest in neural network capability to exhibit "intelligent" behavior

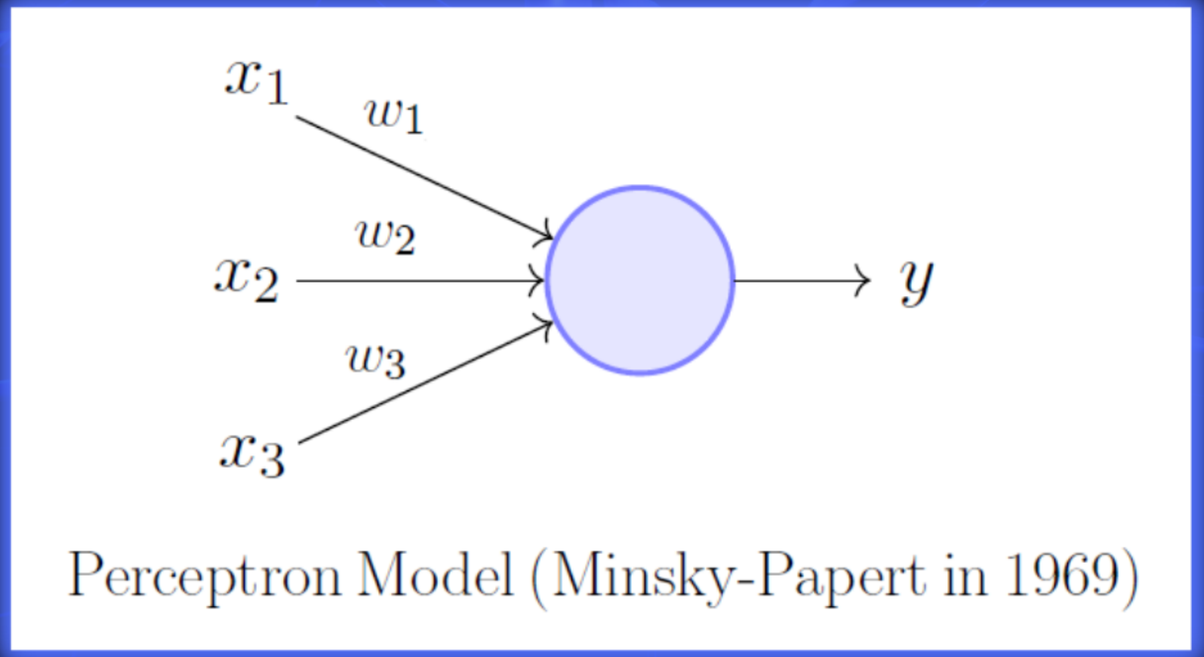
Quick Lingo Catch-Up

$$y = f(x)$$

The goal of supervised machine learning is to find a model *f* that produces a target/output/label *y* given an input *x* by learning from a dataset of *x*-*y* associations.

Perceptrons as Neurons

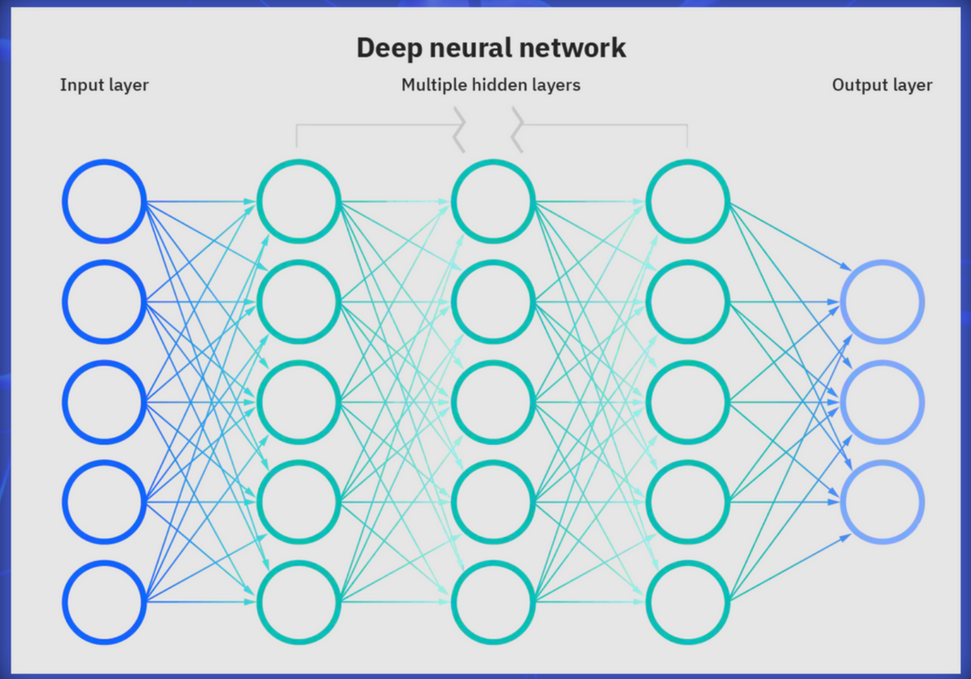
Perceptrons act as a unit of extraction. By stacking them together in intricate patterns, we can obtain sophisticated model output behavior.



Perceptron Model (Minsky-Papert in 1969)

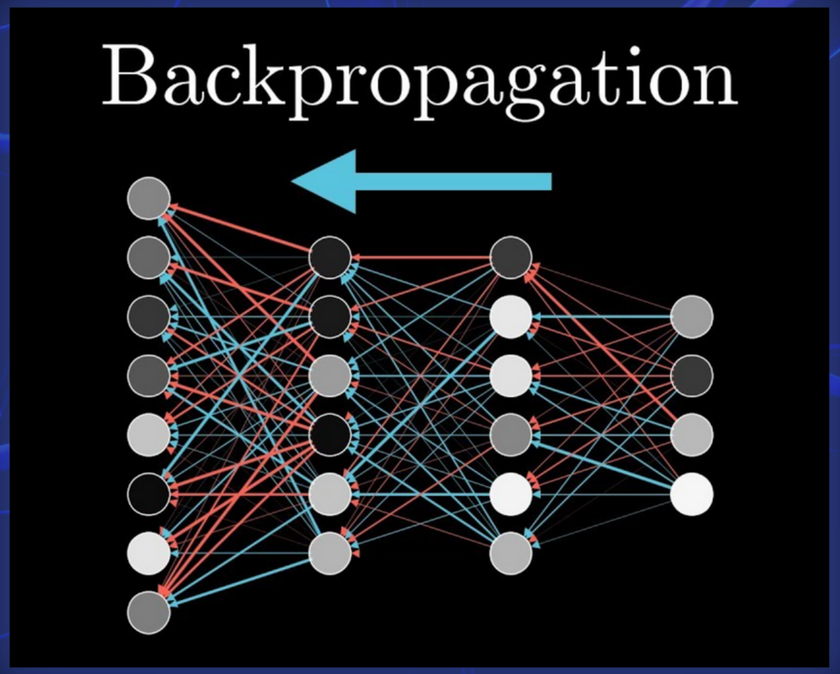
Perceptrons as Neurons

Perceptrons act as a unit of extraction. By stacking them together in intricate patterns, we can obtain sophisticated model output behavior.



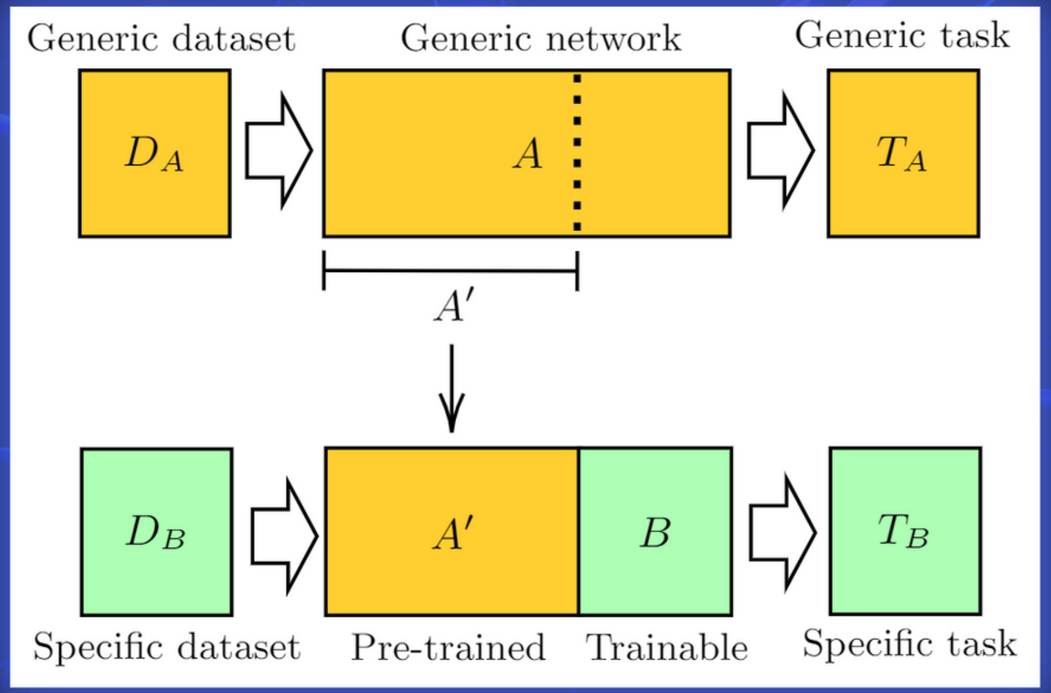
Backpropagation as Learning

Neural networks update internal weights with corresponding direction and magnitude in response to a feedback mechanism.



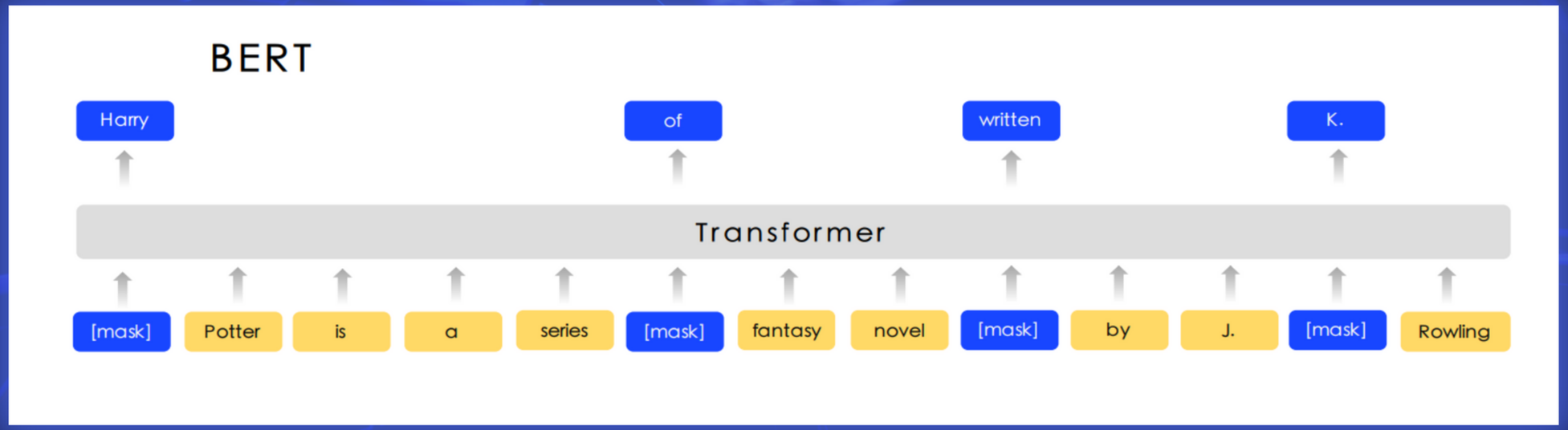
Transfer Learning as Skill Adaptation

Neural networks trained on one standard task can successfully be trained on another specialized task.

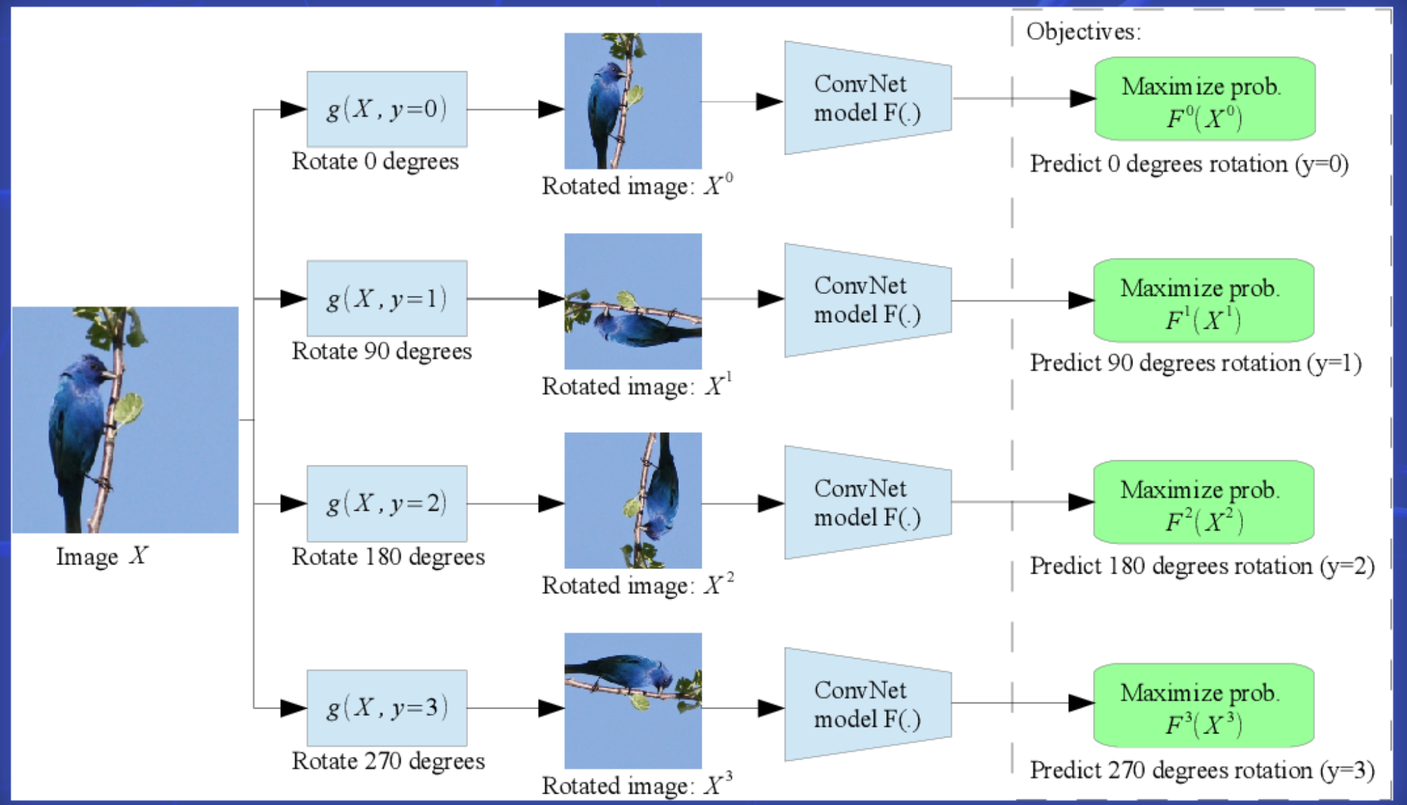


Self-Supervised Learning as Exploration Before Training

The model explores the data landscape without learning labels, then is trained on the specialized task with labels.

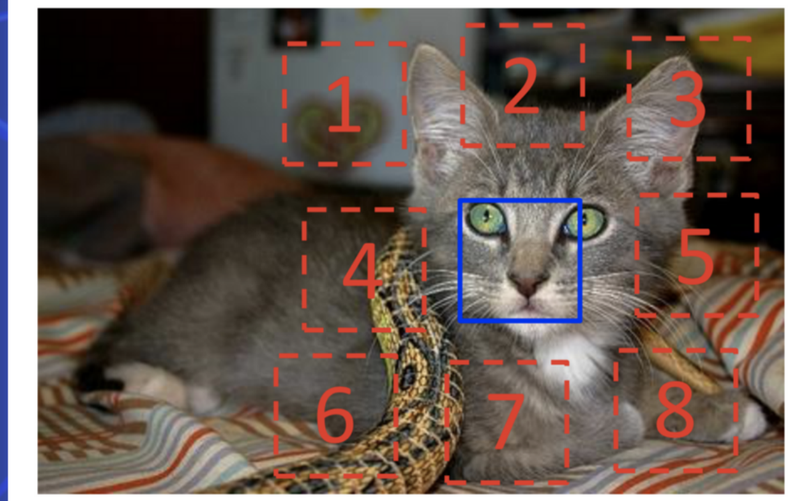


Self-Supervised Learning as Exploration Before Training



https://arxiv.org/abs/1803.07728

Self-Supervised Learning as Exploration Before Training



X = (,); Y = 3

Example:

Question 1:

?

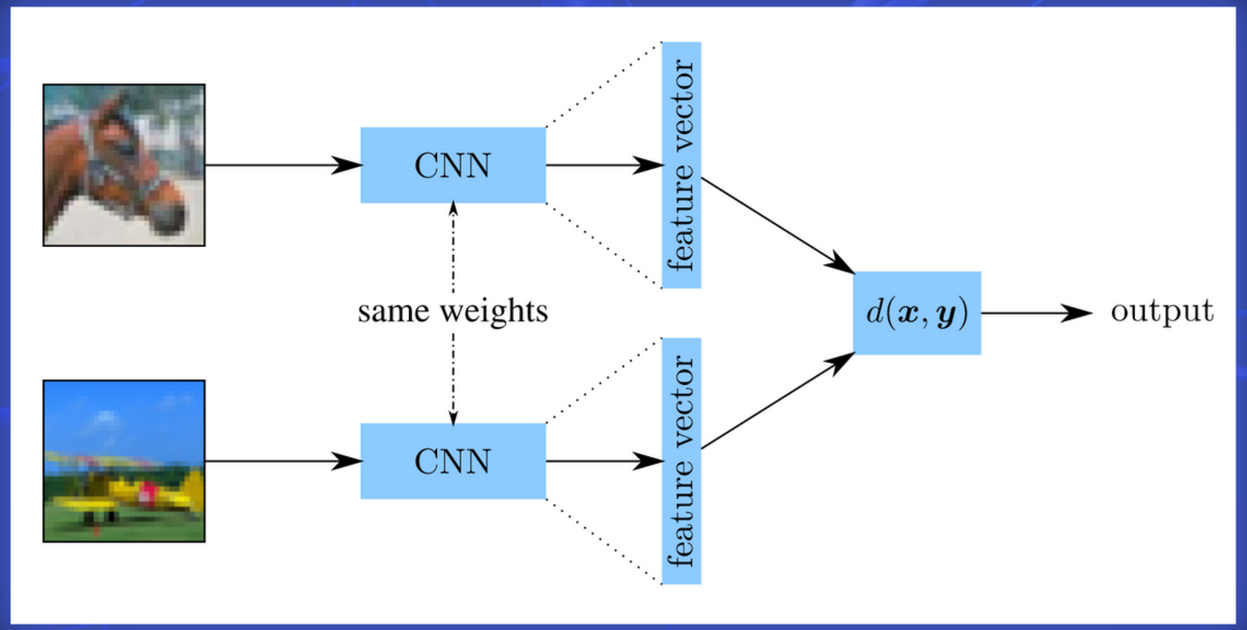
Question 2:

?

https://arxiv.org/abs/1505.05192

Few-Shot Learning as the "Realistic Learning Regime"

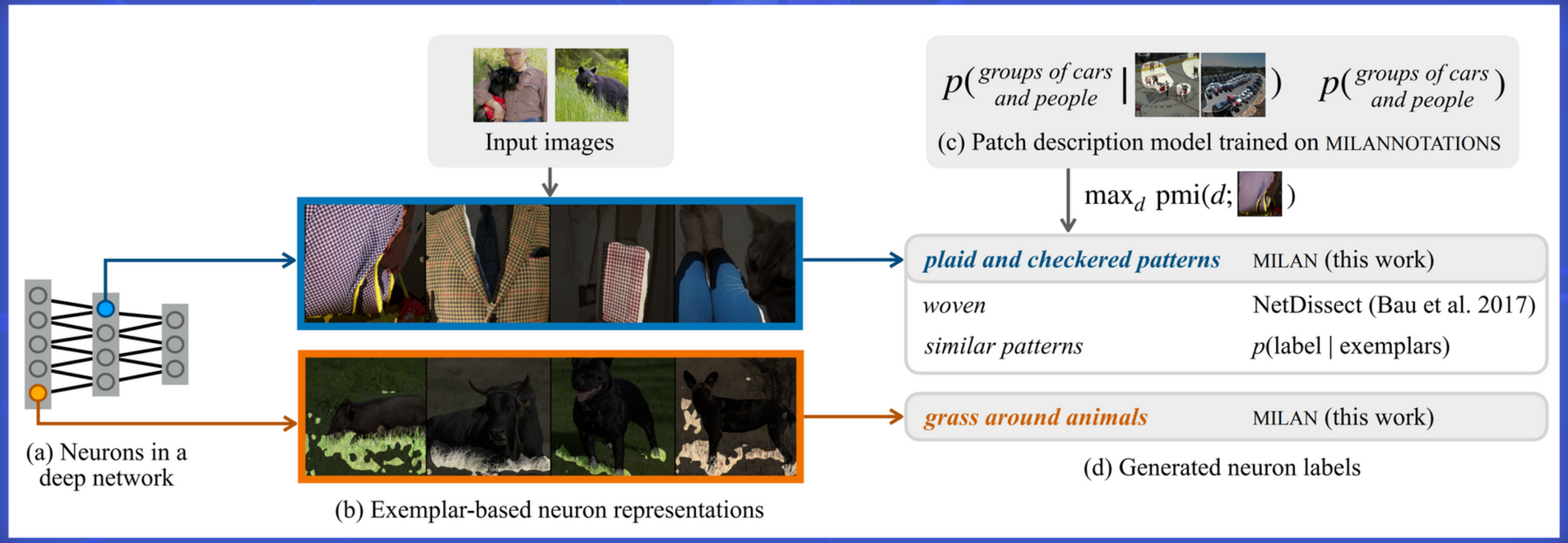
Few-Shot learning models generalize from few samples per class.



https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

NL Weight Interpretation as Decision Explanation

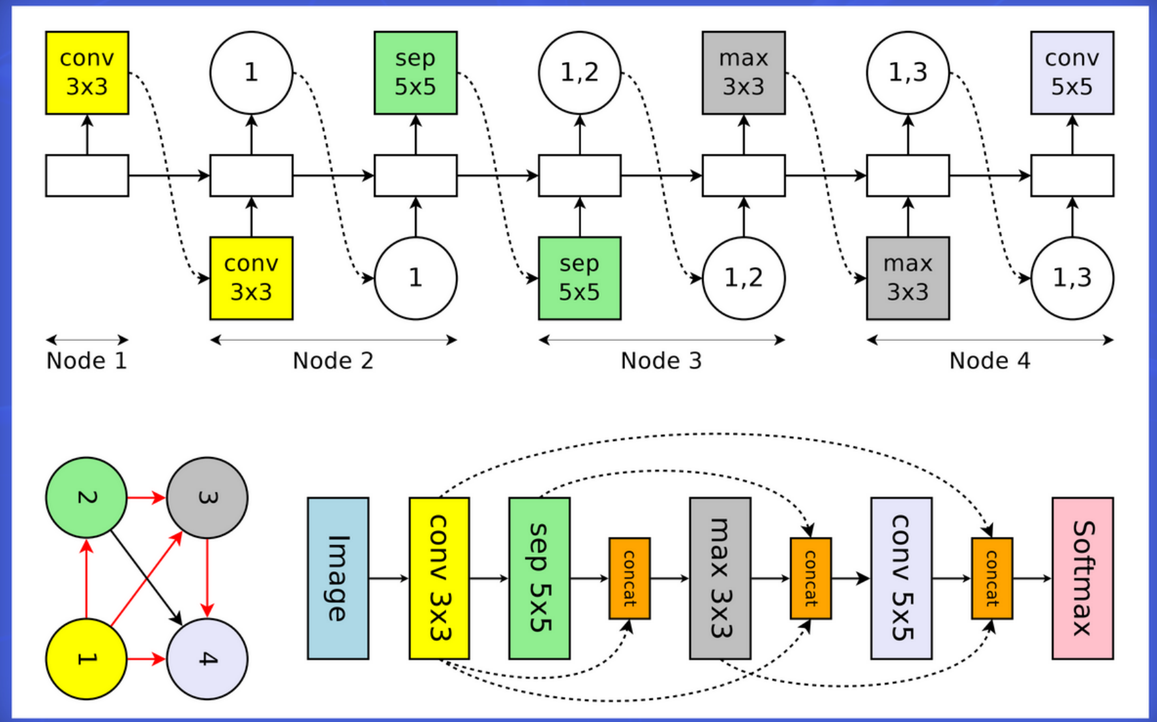
Natural language descr. are discovered from model weights.



https://arxiv.org/pdf/2201.11114.pdf

Neural Architecture Search as Meta-Reflection

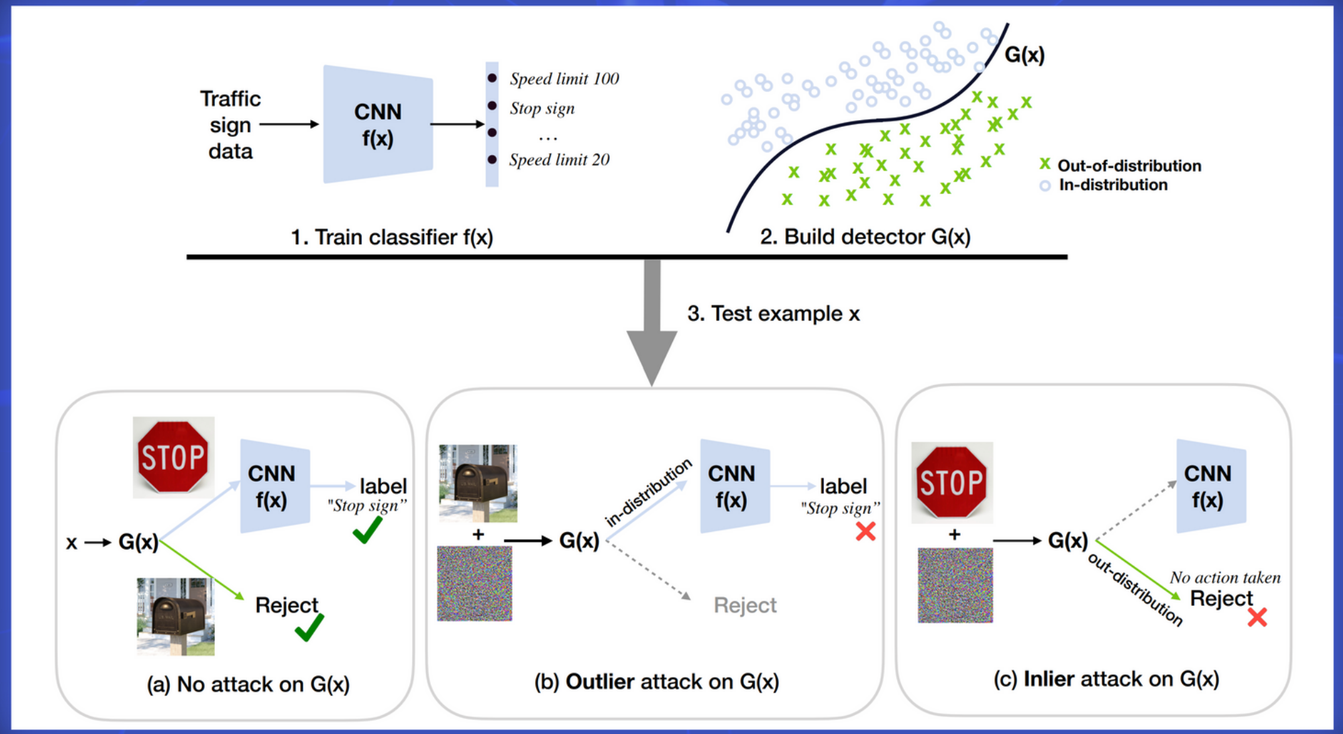
NNs can be used to generate successful NN architectures.



ENAS, https://arxiv.org/pdf/1802.03268.pdf

OOD Detection as Knowing When
You Have No F$(%ing Idea What The Answer Is

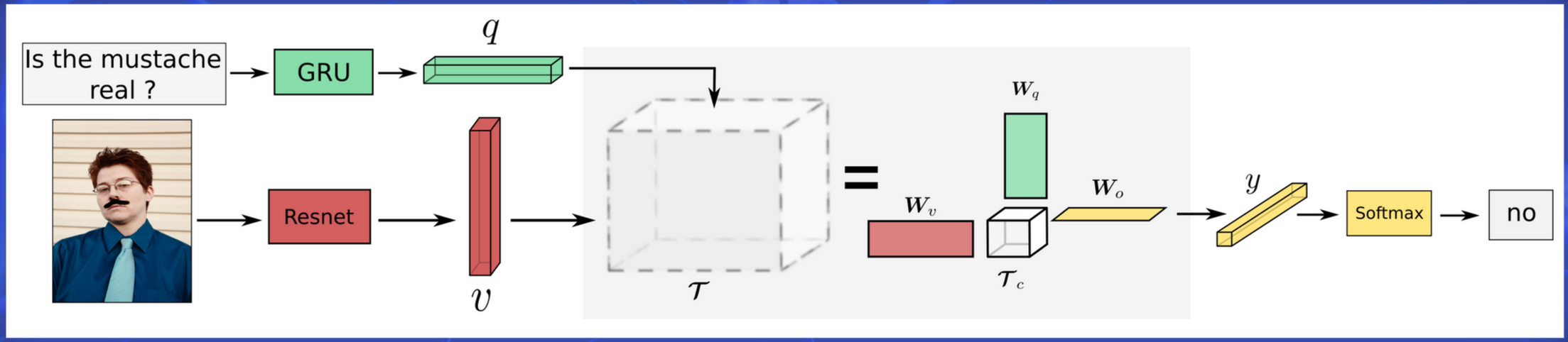
Out of Distribution (OOD) detection allows a system to be 'self-conscious' of its knowledge representation/prediction limits.



ALOE, https://arxiv.org/pdf/2003.09711.pdf

Multimodal Learning as Human-like Perception & Interaction

Models can take in and output multiple modalities at once.



MUTAN, https://arxiv.org/pdf/1705.06676.pdf

Summary of Discussed Ideas

1. **Perceptrons** as Neurons
2. **Backpropagation** as Learning
3. **Transfer Learning** as Skill Adaptation
4. **Self-Supervised Learning** as Exploring Before Training
5. **Few-Shot Learning** as the "Realistic Learning Regime"
6. **NL Weight Interpretation** as Decision Explanation
7. **Neural Architecture Search** as Meta-Reflection
8. **OOD Detection** as Knowing When You Don't Know
9. **Multimodal Detection** as Human-Like Perception & Interaction