

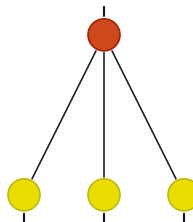
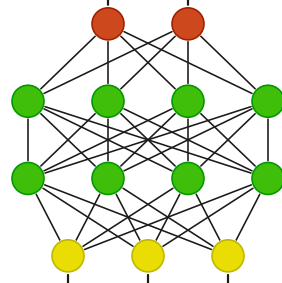
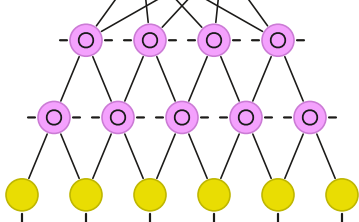
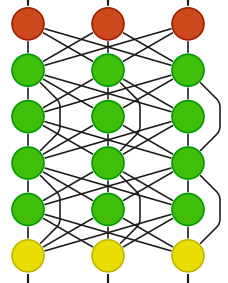
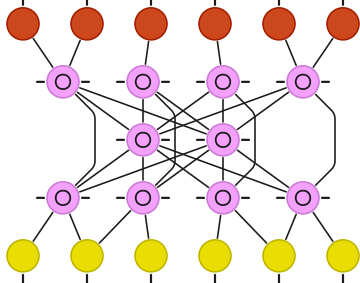
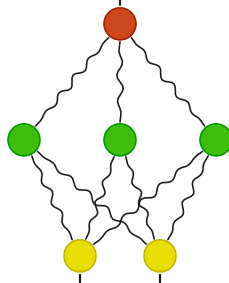
Neural Network Field Guide

42 architectures and training schemes, from the perceptron (1958) to diffusion LLMs (2025).

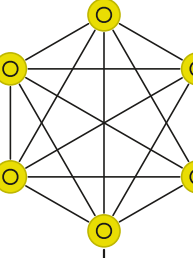
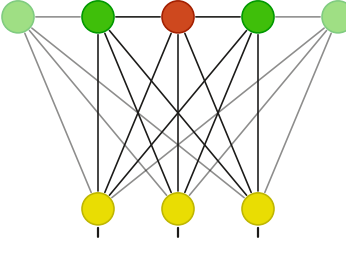
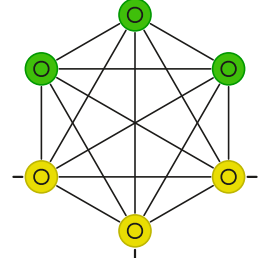
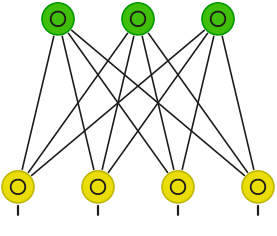
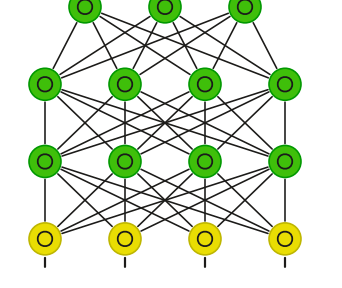
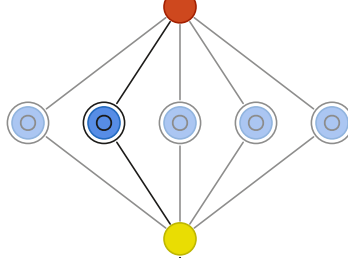
Each diagram draws a network in miniature, from input at the bottom to output at the top. In most panels, a circle is one neuron; in sequence models, where each column is a token, it is that token's whole vector at one layer. A square is larger: an image patch, an expert network, a matrix-valued state. A cell's color and mark both show its role, so the diagrams read with or without color. Faded cells and links take little or no part in the step shown. Links show which tokens can read which. Wherever a part repeats, across positions, windows, copies, or passes, it uses the same weights. Each panel gives a year and the objective or learning rule of its original paper; many can be trained other ways.

Input	Hidden	State	Mixing	Output	Links
Input Image patch Noisy input Random noise Visible unit (in and out) Masked token	Hidden Probabilistic hidden Capsule (vector output) Expert (a whole network) Router	Recurrent Memory cell Gated recurrent Selective state (SSM) Matrix-valued state	Attention Convolution or pooling	Output Reconstruction of input Velocity	Connection Skip connection Data passed on or fed back Gate, acting at the dot Cross-attention, comparison Learnable function

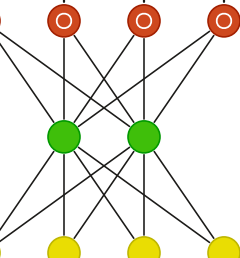
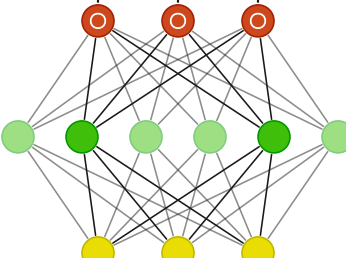
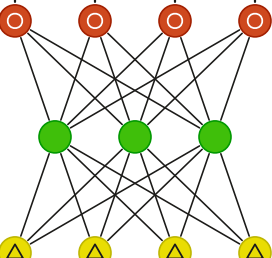
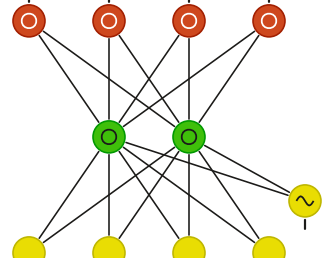
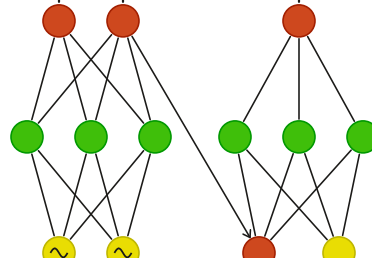
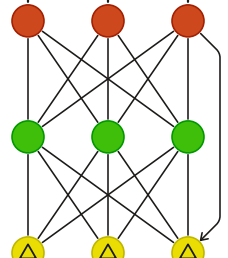
FEED-FORWARD AND CONVOLUTIONAL

Perceptron  Weighted sum, then a threshold 1958 · supervised	MLP  Stacked dense layers 1986 · supervised	CNN  Local filters, shared weights 1989 · supervised	ResNet  Blocks add to a skip path 2015 · supervised	U-Net  Shrink, expand, skip across 2015 · supervised	KAN  Learnable functions on edges 2024 · supervised
---	--	---	--	---	--

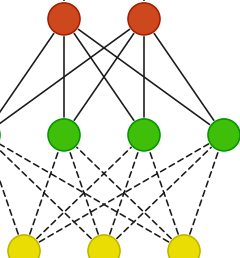
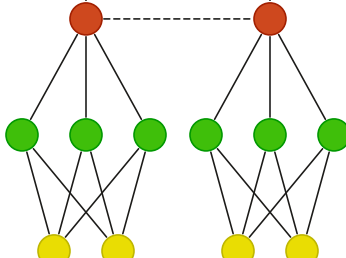
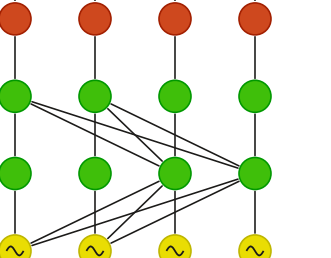
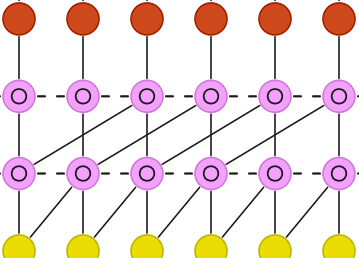
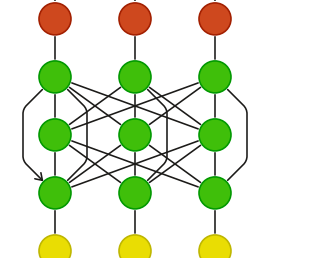
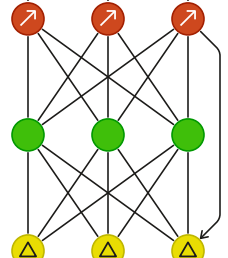
ENERGY-BASED AND SELF-ORGANIZING

Hopfield network  Settles into a stored pattern 1982 · hebbian	Kohonen map  Winner and neighbors move 1982 · competitive	Boltzmann machine  Stochastic, fully connected 1985 · energy	RBM  No links within a layer 1986 · energy	Deep belief network  Stacked RBMs, trained greedily 2006 · energy	Modern Hopfield network  Attention as memory lookup 2020 · energy
---	--	---	---	--	--

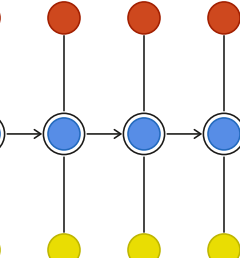
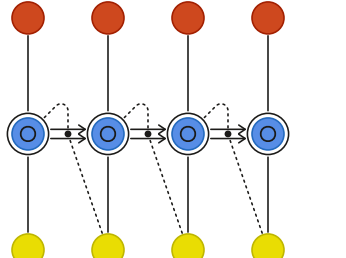
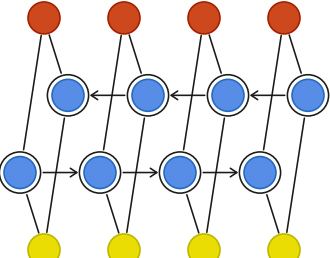
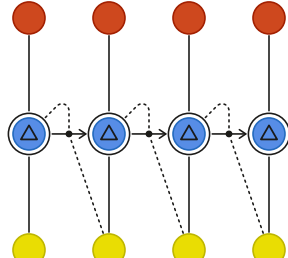
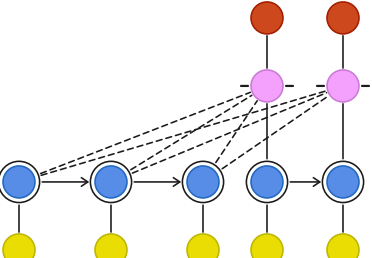
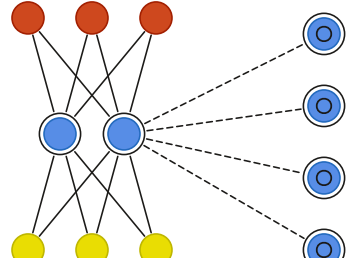
AUTOENCODERS AND GENERATIVE

Autoencoder  Squeeze, then reconstruct 1988 · reconstruct	Sparse autoencoder  Wide code, few active units 2006 · reconstruct	Denoising autoencoder  Clean up a noisy input 2008 · denoise	Variational autoencoder  Sample a code, then decode 2013 · variational	GAN  Generator vs. discriminator 2014 · adversarial	Diffusion model  One denoiser, run T times 2015 · denoise
--	---	---	---	--	--

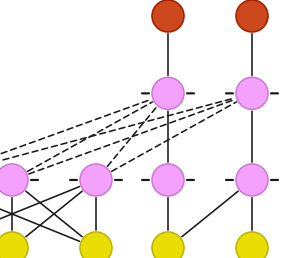
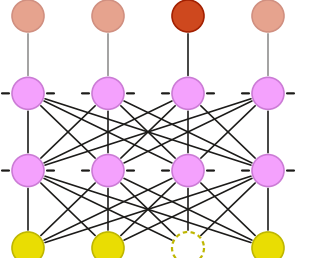
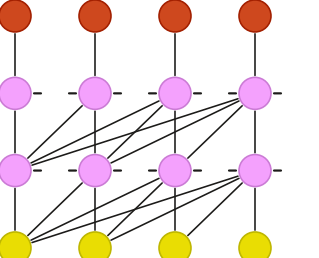
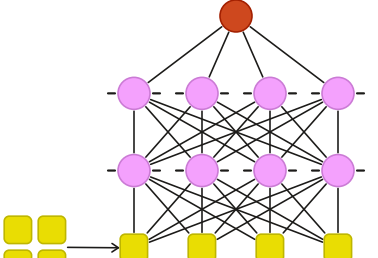
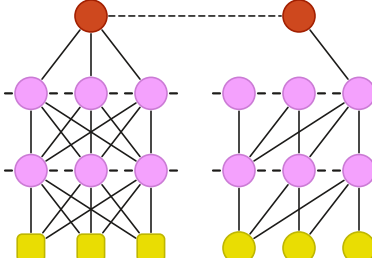
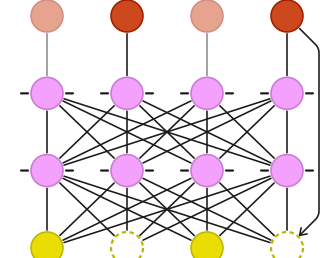
SIMILARITY

Radial basis network  Distance to stored centers 1988 · supervised	Siamese network  Twin towers, shared weights 1993 · contrastive	Normalizing flow  Invertible map from noise to data 2014 · likelihood	WaveNet  Dilated causal convolutions 2016 · next token	Neural ODE  A ResNet with continuous depth 2018 · supervised	Flow matching  Velocity along straight paths 2022 · flow matching
---	--	--	--	---	--

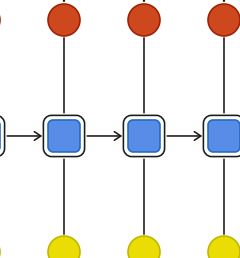
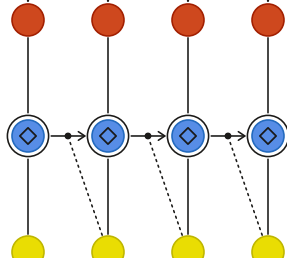
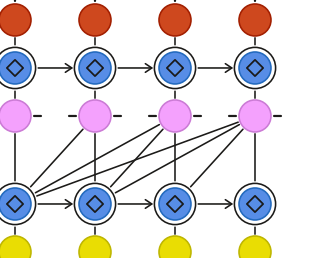
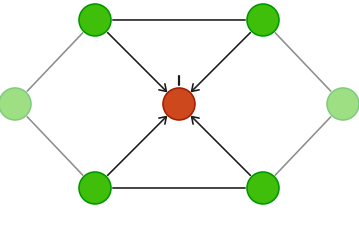
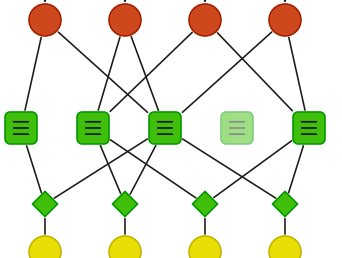
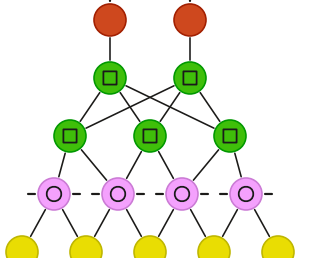
RECURRENT AND MEMORY

RNN  Hidden state passed along 1990 · next token	LSTM  Two states; forget gate from 2000 1997 · supervised	Bidirectional RNN  Reads the sequence both ways 1997 · supervised	GRU  One state, two gates 2014 · next token	Seq2seq with attention  Learns to align and translate 2014 · next token	Neural Turing machine  Controller reads, writes memory 2014 · supervised
---	--	--	--	--	---

ATTENTION

Transformer  Encoder, decoder, cross-attention 2017 · next token	BERT  Every token sees every token 2018 · masked token	GPT-style LLM  Each token sees only the past 2018 · next token	ViT  Image patches become tokens 2020 · supervised	CLIP  Match images to captions 2021 · contrastive	Masked diffusion LLM  Unmask tokens in parallel 2025 · masked token
---	---	---	--	--	--

LINEAR-TIME SEQUENCE MODELS

Linear attention  Attention rewritten as an RNN 2020 · next token	Mamba  Selective state, linear time 2023 · next token	Mamba hybrid  Mostly Mamba, some attention 2024 · next token	Graph neural network  Nodes pool neighbor messages 2005 · supervised	Sparse MoE  Each token → its top-k experts 2017 · next token	Capsule network  Vectors routed by agreement 2017 · supervised
--	--	---	--	---	---