History of Deep-Learning

1.Foundation

1943 — McCulloch & Pitts
- Proposed the first mathematical model of an artificial neuron ("McCulloch–Pitts neuron") and showed networks can represent logical functions.

1949 — Donald Hebb
- "Hebbian learning": idea that synaptic strength increases when neurons fire together ("cells that fire together wire together").

1957–1958 — Frank Rosenblatt
- Invented the Perceptron (early trainable neural classifier). (Commonly referenced across 1957 report + 1958 publications.)

2. First boom → first winter (1960–1979)

1960s (range) — Early neural hardware + adaptive filters
- Important practical work happened (e.g., ADALINE/MADALINE era), but the bigger "story point" is that single-layer models dominated.

1969 — Minsky & Papert
- Book Perceptrons highlighted limitations of single-layer perceptrons (famously XOR-like limits), contributing to reduced funding/interest for neural nets for a while.

1974 — Paul Werbos
- Described what we now call backpropagation (reverse-mode differentiation) in his thesis, a key stepping stone for training multi-layer nets.

3.Backprop era + modern building blocks (1980-1999)

1980 — Kunihiko Fukushima
- Neocognitron: an early ancestor of CNNs (hierarchical feature extraction; shift-invariant recognition).

1982 — John Hopfield
- Hopfield networks: recurrent associative memory; linked neural nets with energy-based formulations.

1986 — Rumelhart, Hinton, Williams
- Popularized backpropagation for multi-layer networks in a highly influential Nature paper.

1989 — Universal Approximation (Cybenko)
- Proved (under conditions) that a neural network with 1 hidden layer can approximate continuous functions (important theory milestone).

1989 & 1998 — Yann LeCun and colleagues
- 1989: backprop applied to real-world handwriting / zip code recognition (CNN-style systems emerging).
- 1998: "Gradient-Based Learning Applied to Document Recognition" (LeNet-style CNNs) formalized CNN practice for vision.

1997 — Hochreiter & Schmidhuber
- LSTM introduced to solve long-term dependency + vanishing gradient issues in RNNs.

4. "Deep Learning" revival (2000–2011)

2006 — Hinton, Osindero, Teh
- Deep Belief Nets (DBNs) + greedy layer-wise pretraining helped revive training deeper models.

2010 — Nair & Hinton
- Popularized ReLU benefits (Rectified Linear Units) for training efficiency / performance.

5. The big breakthrough + scaling era (2012–2016)

2012 — Krizhevsky, Sutskever, Hinton (AlexNet)
- **CNN + GPUs + big data (ImageNet) triggered the modern deep learning boom in computer vision.**

2013 — Mikolov et al.
- **word2vec made neural word embeddings mainstream and hugely boosted NLP pipelines.**

2014 — Dropout (Srivastava et al.)
- **Regularization technique that improved generalization for large nets.**

2015 — Batch Normalization (Ioffe & Szegedy)
- **Stabilized and sped up deep net training by normalizing activations.**

2015 — ResNet (He et al.)
- **Residual connections enabled very deep networks (100+ layers) reliably.**

2016 — AlphaGo (Silver et al.)
- **Landmark: deep neural networks + tree search + RL beat top human Go players.**

7.Transformer & Foundation Model Era

2017 — Vaswani et al. (Transformer)
- **Attention-only architecture replaced RNNs/CNNs and enabled scalable sequence modeling.**

2018 — Devlin et al. (BERT)
- **Bidirectional transformer pretraining improved contextual language understanding.**

2018 — Radford et al. (GPT)
- **Generative pretraining showed strong transfer learning for NLP tasks.**

2019 — He et al. (EfficientNet)
- **Compound scaling balanced depth, width, and resolution efficiently.**

2020 — Brown et al. (GPT-3)
- **Large-scale transformers demonstrated few-shot and zero-shot learning.**

2020 — Jumper et al. (AlphaFold2)
- **Deep learning solved protein structure prediction with near-experimental accuracy.**

2021 — Dosovitskiy et al. (Vision Transformer)
- **Transformers successfully replaced CNNs in computer vision.**

2022 — ChatGPT (Instruction-tuned LLMs)
- **RLHF + transformers enabled human-aligned conversational deep learning models.**

2023 — Multimodal Foundation Models
- **Single deep models jointly learned text, image, audio, and video representations.**

2024 — Diffusion Transformers & Video Models
- **Diffusion-based deep networks achieved high-quality image and video generation.**

2025 — Reasoning-focused Deep Models
- **Deep learning shifted toward structured reasoning, efficiency, and tool-using models.**