Deep learning is a subset of machine learning that uses layered artificial neural networks, inspired by the human brain, to learn from large amounts of data. It excels at recognizing complex patterns in unstructured data, like images, text, and sounds, enabling it to power applications such as self-driving cars, facial recognition, and speech recognition. The "deep" in deep learning refers to the multiple layers of the neural network, which allow the system to automatically learn and extract increasingly complex features from the data.

How it works

- **Artificial neural networks**:

  Deep learning models use networks of interconnected "neurons" organized in layers, similar to how neurons are connected in the human brain.

- **Hierarchical feature learning**:

  Each layer in the network processes the output from the previous one, learning to recognize progressively more complex features. For example, an early layer might identify simple shapes, while a later layer uses that information to recognize a specific object like a stop sign.

- **Training and optimization**:
  The network "learns" by adjusting the strength of the connections between neurons through a process of trial and error. It uses this feedback to improve its accuracy over time.

Applications

- **Image and speech recognition**:

  Used in systems like facial recognition, voice assistants (e.g., Siri, Alexa), and systems that can describe images or transcribe speech to text.

- **Natural language processing**:

  Powers applications like language translation and predictive text features.

- **Autonomous systems**:

  A core technology for self-driving cars, enabling them to recognize objects, pedestrians, and other vehicles.

- **Healthcare**:

  Assists in tasks like analyzing medical images to detect tumors or other indicators of disease.

- **Other areas**:

Also used in fraud detection, cybersecurity, and business intelligence.