Технологията за компютърно виждане е една от най-обещаващите области на научни изследвания в рамките на изкуственият интелект и компютърните науки, и предлага огромни предимства за бизнеса в нашето съвремие. В сърцевината си, областта на компютърното зрение се фокусира върху проектиране на компютърни системи, които притежават способността да улавят, да разбират, и да интерпретират важни визуална информация, която се съдържа в изображенията и видео данните. След това компютърните визуални системи превеждат тези данни, използвайки контекстуални знания, предоставени от хора, в прозрения, използвани за задвижване на вземането на решения. Превръщането на едно изображение от суров вид в концепция от по-високо ниво, така че хората или компютрите да могат да интерпретират и действат по него, е основната цел на технологията за компютърното виждане. Трябва обаче, да се направи важно разграничение между компютърното зрение и обработката на изображението. Обработката на изображения е науката за извършване на промени на един образ, така че да произведе нов образ с определени засилени характеристики. Тези промени включват – увеличаването на резолюцията, нормализиране на яркостта и контраста, изрязване, замъгляване, или някаква друга дигитална трансформация, която ще е необходима за конкретната цел. Дигиталната обработка на изображения не взема предвид действителното съдържание на изображението - това е просто поредица от механични трансформации, предприети да променят изображението за някаква определена цел.

Системите за човешкото зрение имат огромното предимство да бъдат информирани за знания за цял живот, които помагат да се контекстуализират данните в нашето зрително поле. Нашите очни ябълки улавят визуалната информация например - образа на една котка, и нашият предишен опит интерпретира това събиране на отразена светлина и се свързва с понятието за котка. Сложността на нашата система за визуално възприятие и тясната ѝ връзка с нашата памет и по-високи способности за разсъждение дава на тези визуални данни необходимият контекст, за да осигури качество в ежедневните активности. Тези човешки способности, докато са недостъпни за компютрите, могат да бъдат имитирани ефективно чрез алгоритми за машинно обучение. Но както се оказва, обучаването на машини, които да имитират тази основна човешка фунцкия, която е гордо демонстрирана от 5 годишни деца по цял свят, е изключително трудно. Решаването на този проблем непрекъснато заема най-ярките умове в изследванията на изкуственият интелект.

Дълбокото обучение или йерархичното обучение е подмножество на машинното обучение в изкуственият интелект, който може да имитира функцията за обработка на данни на човешкия мозък и да създава подобни модели, които мозъкът използва за вземане на решения. Противно на алгоритмите, базирани на задачи, системите за дълбоко обучение се учат от представянето на данни - те могат да се учат от неструктурирани или немаркирани данни. Архитектурите за дълбоко обучение като дълбоките невронни мрежи, мрежи от вярвания, повтарящи се невронни мрежи, и конволюционните невронни мрежи са намерили приложения в областта на компютърното зрение, аудио/говорното разпознаване, машинният превод, филтрирането в социалните мрежи, биоинформатиката, дизайнът за лекарства и много други.

Невронните мрежи са съставени от набор от алгоритми, които са моделирани по човешкият мозък. Тези алгоритми могат да интерпретират сензорни данни чрез машинно възприятие и надписвайки или групирайки необработените данни. Те са предназначени да разпознават числови модели, които се съдържат във вектори, в които са всички данни от реалния свят (изображения, звуци, текстове, времеви редове и т.н) трябва да бъдат преведени. В основата си, основната задача на невронните мрежи е да групира и класифицира необработените данни - те групират немаркираните данни въз основа на приликите, които са намерени във входните данни и след това се класифицират данните въз основа на етикетираният набор от данни за обучение. Невронните мрежи могат автоматично да се адаптират към променящият се вход. Така че, не е необходимо наново да проектираме изходните критерии всеки път, когато входът се промени, за да генерира възможно най-добрия резултат.